Universidad Carlos III de Madrid
www.uc3m.es

TESIS DOCTORAL

Essays on the Identification of Linear Rational Expectations Models

Autor:

Mehdi Hamidisahneh

Director de Tesis:

Carlos Velasco

Departamento de Economia

Getafe, Junio 2017

Universidad Carlos III de Madrid
www.uc3m.es

(a entregar en la Oficina de Posgrado, una vez nombrado el Tribunal evaluador , para preparar el documento para la defensa de la tesis)

TESIS DOCTORAL

Essays on the Identification of Linear Rational Expectations Models

Autor:
Mehdi Hamidisahneh

Director De Tesis:
Carlos Velasco

Firma del Tribunal Calificador:

Firma

Presidente: (Nombre y apellidos)

Vocal: (Nombre y apellidos)

Secretario: (Nombre y apellidos)

Calificación:

Getafe, 8 de Junio de 2017

Contents

Are Small-Scale SVARs Useful for Business Cycle Analysis? Revisiting Non-Fundamentalness

Fabio Canova
BI Norwegian Business School, and CEPR *

Mehdi Hamidi Sahneh
Universidad Carlos III †

June 6, 2017

Abstract

Non-fundamentalness arises when current and past values of the observables do not contain enough information to recover SVAR disturbances. Using Granger causality tests, the literature suggested that several small scale SVAR models are non-fundamental and thus not necessarily useful for business cycle analysis. We show that causality tests are problematic when SVAR variables cross sectionally aggregate the variables of the underlying economy or proxy for non-observables. We provide an alternative testing procedure, illustrate its properties with Monte Carlo simulations, and re-examine a prototypical small scale SVAR model.

Keywords: Aggregation; Non-Fundamentalness; Granger causality, Small scale SVARs.

JEL classification: C5, C32, E5.

---

*Centre for Applied Macro and Petroleum Economics and Department of Economics, BI Norvegian Business School, Nydalsveien 37, 0484 Oslo, Norway.

†Departamento de Economia, Universidad Carlos III de Madrid, Getafe, 28903, Spain. We thank the editor (Claudio Michelacci), five anonymous referees, Carlos Velasco, Jesus Gonzalo, Hernan Seoane, Gianni Amisano, Jordi Gali, Davide de Bortoli, Luca Sala, Benjamin Holcblat, Domenico Giannone, Lutz Kilian, Luca Gambetti and Mario Forni and the participants of seminars at BI Norwegian Business School, UPF, Bocconi, Humboldt, Luiss, University of Glasgow, University of Helsinki, and Federal Reserve Bank of New York, Federal Reserve Board, and of the 2016 IAAE and 2016 ESEM conferences for comments and discussions. Canova acknowledges the financial support from the Spanish Ministerio de Economia y Competitividad through the grants ECO2012-33247 and ECO2015-68136-P and FEDER, UE.

1 Introduction

Structural Vector Autoregressive (SVAR) models have been extensively used over the last 30 years to study sources cyclical fluctuations . The methodology hinges on the assumption that structural shocks can be obtained from linear combinations of current and past values of the observables. Non-fundamentalness arises when this is not the case. In a non-fundamental system, structural shocks obtained via standard identification procedures may have little to do with the true disturbances, even when identification is correctly performed, making SVAR evidence unreliable.

Since likelihood or spectral estimation procedures can not distinguish fundamental vs. non-fundamental Gaussian systems (see e.g. Canova (2007), page 114), it is conventional in applied work to rule out all the non-fundamental representations that possess the same second-order structure of the data. However, this choice is arbitrary. There are rational expectation models (Hansen and Sargent, 1991), optimal prediction models (Hansen and Hodrick, 1980), permanent income models (Fernndez-Villaverde et al., 2007), news shocks models (Forni et al., 2014), and fiscal foresight models (Leeper et al., 2013), where optimal decisions may generate non-fundamental solutions. In addition, non-observability of certain states or particular choices of observables may make fundamental systems non-fundamental.

Despite the far-reaching implications it has for applied work, little is known on how to empirically detect non-fundamentalness. Following the lead of Lütkepohl (1991), Giannone and Reichlin (2006) and Forni and Gambetti (2014) (henceforth, FG) suggest that, under fundamentalness, external information should not Granger cause VAR variables. Using such a methodology, FG and Forni et al. (2014) argued that several small scale SVARs are non-fundamental, thus implicitly questioning the economic conclusions that are obtained. Considering the popularity of small scale SVARs in macroeconomics, this result is disturbing. This paper shows that Granger causality diagnostics may lead to spurious results in common and relevant situations.

Why are there problems? Because of small samples, instabilities, identification or interpretation difficulties, one typically uses a small scale SVAR to examine the transmission of relevant disturbances, even if the process generating the data (DGP) features many more variables and shocks. But the shocks recovered by such SVAR systems are linear combinations of a potentially larger set of primitive structural shocks driving the economy. Thus, any variable excluded from the SVAR, but containing information about these primitive disturbances, predicts SVAR shocks (and thus Granger cause the endogenous variables), regardless of whether the model is fundamental or not.

To illustrate the point, suppose we want to measure the effects of technology shocks on economic activity. Small scale SVARs designed for this purpose typically include an aggregate measure of labour productivity, hours, and a few other aggregate variables. Suppose that what drives the economy are sector-specific, serially correlated productivity disturbances. The technology shock recovered from an SVAR will be a linear transformation of current and past sectoral productivity shocks. Since, e.g., sectoral capital or sectoral labour productivity have information about sectoral disturbances, they will predict SVAR technology shocks, both when the model is fundamental and when it is not.

A similar problem occurs when the SVAR features a proxy variable. For example, TFP is latent and typical estimates are obtained from output, capital and hours worked data. If capital and hours worked are excluded from the SVAR, any variable that predicts them will Granger cause estimated TFP, regardless of whether the model is fundamental or not.

In general, whenever a small scale SVAR is used, aggregation rather than non-fundamentalness may be the reason for why Granger causality tests find predictability. Thus, if non-fundamentalness is of interest, it is crucial to have a testing approach which is robust to aggregation and non-observability problems. We propose an alternative procedure, based on ideas of Sims (1972), which is has this property and exploits the fact that, under non-fundamentalness, future SVAR shocks predict a vector of variables excluded from the SVAR.

We perform Monte Carlo simulations using a version of the model of Leeper et al. (2013) as DGP with capital tax, income tax, and productivity disturbances. We assume that the

SVAR includes capital and an aggregate tax variable (or an aggregate tax rate computed from revenues and output data) and show that our approach has good small sample properties. In contrast, spurious non-fundamentalness arises with standard diagnostics. Absent aggregation problems, our approach and Granger causality tests have similar small sample properties.

We re-examine the small scale SVAR employed by Beaudry and Portier (2006) designed to measure the macroeconomic effects of news. We find that the model is fundamental according to our test but non-fundamental according to a Granger causality diagnostic. We show that the rejection of the null with the latter is due to aggregation: once coarsely disaggregated TFP data is used in the SVAR, Granger causality no longer rejects the null of fundamentalness.

The dynamics responses to news shocks in the systems with aggregated and disaggregated TFP measures are however similar (see also Beaudry et al. (2015)). Thus, the SVAR disturbances the two systems recover are likely to be similar combinations of the primitive structural shocks and, thus, not necessarily economically interpretable.

Two caveats have to be mentioned. First, it is well-known that non-fundamental representations are identifiable under some non-Gausianity assumption. See for instance Lii and Rosenblatt (1982) and Hamidi Sahneh (2014). However, checking whether a Gaussian VAR is fundamental or not is complicated because the likelihood function or the spectral density can not distinguish between a fundamental and a non-fundamental representations. Importantly, our analysis allows for, but not restricted to the Gaussian shocks. Second, although we focus on SVARs, our procedure also works for SVARMA models, as long as the largest MA root is sufficiently away from unity.

The rest of the paper is organized as follows. Section 2 provides examples of non-fundamental systems and highlights the reasons for why problem occurs. Section 3 shows why standard tests may fail and propose an alternative approach. Section 4 examines the performance of various procedures using Monte Carlo simulations. Section 5 investigates the properties of a small scale SVAR system. Section 6 concludes.

2 A few example of non-fundamental systems

As Kilian and Lütkepohl (2016) highlighted, the literature has primarily focused on non-fundamentalness driven by a mismatch between agents and econometricians information sets, because of omitted variables (see e.g. Giannone and Reichlin (2006), Kilian and Murphy (2014)), or of the timing of news revelation (see e.g. Leeper et al. (2013) , Forni et al. (2014)). However, there may be other reasons for why it emerges.

First, non-fundamentalness may be intrinsic to the optimization process and to the modelling choices an investigator makes, see e.g. Hansen and Sargent (1980, 1991). Optimizing models producing non-fundamental solutions are numerous; the next example shows one.

Example 1. Suppose the dividend process is $d_t = e_t - a\mathrm{e}_{t-1}$, where $a < 1$, and suppose stock prices are expected discounted future dividends: $p_t = E_t \sum_j \beta^j d_{t+j}$, $0 < \beta < 1$. The equilibrium value of p_t in terms of the dividends innovations is

$$p_t = (1 - \beta a)e_t - ae_{t-1} \tag{2.1}$$

Thus, even though the dividends process is fundamental ($a < 1$), the process for stock prices could be non-fundamental if $|\frac{(1-\beta a)}{a}| < 1$, which occurs when $\frac{1}{1+\beta} < a$. If $a \geq 0.5$, any economically reasonable value of β will make stock prices non-fundamental. On the other hand, if we allow stock prices to have a bubble component e_t^b whose expected value is zero, the vector (e_t, e_t^b) is fundamental for (d_t, p_t), regardless of the value of β. Thus, allowing for bubbles in theory makes a difference as far as recovering dividend shocks from the data. □

Second, non-fundamentalness may be due to non-observability of some of the endogenous variables of a fundamental model. The next example illustrates how this is possible.

Example 2. Suppose the production function (in logs) is:

$$Y_t = K_t + e_t \tag{2.2}$$

and the law of motion of capital is:

$$K_t = (1-\delta)K_{t-1} + ae_t \tag{2.3}$$

If both (K_t, Y_t) are observable this is just a bivariate restricted VAR(1) and e_t is fundamental for both (k_t, y_t). However, if the capital stock is unobservable, (2.2) becomes

$$Y_t - (1-\delta)Y_{t-1} = (1+a)e_t + (1-\delta)e_{t-1} \tag{2.4}$$

Clearly, if $a < 0$ and $|a| < |\delta|$, e_t can not be expressed as a convergent sum of current and past values of Y_t and (2.4) is non-fundamental. In addition, if δ and a are both small, (2.4) has a MA root close to unity and a finite order VAR for Y_t poorly approximates the underlying bivariate process; see also Ravenna (2007), and Giacomini (2013).

Third, a particular variable selection may induce non-fundamentalness, even if the system is, in theory, fundamental. Hansen and Hodrick (1980) showed that this happens when forecast errors are used in a VAR. The next example shows a less known situation.

Example 3. Consider a standard consumption-saving problem. Let income $Y_t = e_t$ be a white noise. Let $\beta = \frac{1}{R} < 1$ be the discount factor and assume quadratic preferences. Then:

$$C_t = C_{t-1} + (1 - R^{-1})e_t \tag{2.5}$$

Thus, growth rate of consumption has a fundamental representation. However, if we setup the empirical model in terms of savings, $S_t \equiv Y_t - C_t$, the solution is

$$S_t - S_{t-1} = R^{-1}e_t - e_{t-1} \tag{2.6}$$

and the growth rate of saving is non-fundamental. □

In sum, there may be many reasons for why an empirical model may be non-fundamental.

Assuming away non-fundamentalness is problematic. Focusing on omitted variable or anticipation problems is, on the other hand, reductive. One ought to have procedures able to detect whether a SVAR is fundamental and, if it is not, whether violations are intrinsic to theory or due to applied investigators choices.

3 The Setup

Because in this section we need to distinguish the structural disturbances driving the fluctuations in the DGP from the shocks a SVAR may recover, we use the convention that " primitive" structural shocks are the disturbances of the DGP and "SVAR" structural shocks those obtained with the empirical model.

We assume that the DGP for the observables can be represented by an n-dimensional vector of stationary variables χ_t driven by $s \geq n$ serial and mutually uncorrelated primitive structural shocks ς_t.

Assumption 1. (Macroeconomic representation) The vector χ_t satisfies

$$\chi_t = \Gamma(L)\varsigma_t$$

where $\Gamma(L) = \sum_{i=0}^{\bar{Q}} \Gamma_i L^i$, $\Gamma_0 = I$, Γ_i's are $(n \times n)$ matrices each i, L is the lag operator, and $\sum_{i=0}^{\bar{Q}} \Gamma_i^2 < \infty$.

The DGP in (3) is quite general and covers, for example, stationary dynamics general equilibrium (DSGE) models solved around a deterministic steady state or non-stationary DSGEs solved around a deterministic or a stochastic balanced growth path. Assumption 1 places mild restrictions on the roots of $\Gamma(L)$. In theory, ς_t could be fundamental for χ_t or not.

Given a typical sample, n the dimension of χ_t is generally large and $\Gamma(L)$ is of infinite dimension. Thus, for estimation and inferential purposes an applied investigator typically confines attention to an m-dimensional vector x_t, where $\mathcal{H}_t^x \subset \mathcal{H}_t^\chi$, and $\mathcal{H}_t^j$ is the closed linear span of $\{j_s : s \leq t\}, j_t = (x_t, \chi_t)$ [1].

Assumption 2. (VAR information set) The vector x_t is driven by a $m \times 1$ vector of mutually and serially uncorrelated SVAR structural shocks $\varsigma_{x,t} \sim iid(0, \sigma_\varsigma)$:

$$
\begin{aligned}
x_t &= \Gamma_x(L)\varsigma_t && (3.1)\\
&\equiv \Pi(L)u_t && (3.2)
\end{aligned}
$$

where $\Gamma_x(L)(m < n)$ is an $m \times m$ matrix of lag polynomial, $\Pi(L) = \sum_{i=0}^{Q \leq \bar{Q}} \Pi_i L^i$, Π_i are $m \times m$ matrices for each i, $\Pi_0 = I$, $\sum_{i=0}^{Q \leq \bar{Q}} \Pi_i^2 < \infty$, and u_t is a white noise process obtained from linear transformation of current and past primitive structural shocks.

Equation (3.1) covers many cases of interest in macroeconomics. For example, x_t may contain a subset of the variables belonging to χ_t, linear combinations, regression residuals, or forecast errors computed from the elements of χ_t. Thus, the framework includes the case of a variable belonging to the DGP but unobserved and thus omitted from the empirical model (as in example 2); the situation where the DGP has disaggregated variables but the empirical model is set up in terms of aggregated variables; the case where the DGP has an unobservable variable (e.g. total factor productivity) proxied by a linear combination of observables (i.e. output, capital and labor); and the case where all DGP variables are observables (e.g., we have consumption data) but the empirical model contains linear combinations of the observables (i.e. savings as in example 3).

Since the dimension of ς_t is larger than the dimension of x_t, cross-sectional aggregation occurs. That is, the econometrician estimating an SVAR may be able to recover the $m \times 1$ vector $\varsigma_{x,t}$ from the reduced form residuals u_t, but never the $s \times 1$ vector ς_t. For example, the DGP may describe a small open economy subject to external shocks coming from many countries, while the empirical model is specified so that only rest of the world variables are used. If $\Gamma(L)$ has a block exogenous structure, it may be possible to aggregate the vector

[1] The linear span is the smallest closed subspace which contains the subspaces.

external shocks into one shock without contamination from other disturbances, see e.g. Faust and Leeper (1997). However, even in this case, it is clearly impossible to recover the full vector of country specific external disturbances.

Next, we provide the definition of fundamentalness for the empirical model (3.2) (see also Rozanov (1967)) and Alessi et al. (2011)).

Definition 1: An uncorrelated process $\{u_t\}$ is x_t-fundamental if $\mathcal{H}_t^u = \mathcal{H}_t^x$ for all t. It is non-fundamental if $\mathcal{H}_t^u \subset \mathcal{H}_t^x$ and $\mathcal{H}_t^u \neq \mathcal{H}_t^x$, for at least one t.

The empirical model (3.2) is fundamental if and only if all the roots of the determinant of the $\Pi(L)$ polynomial lie outside the unit circle in the complex plane - in this case $\mathcal{H}_t^u = \mathcal{H}_t^x$, for all t. Alternatively, the model is fundamental if it is possible to express u_t as a convergent sum of current and past x_t's. Fundamentalness is closely related to the concept of invertibility: the latter requires that no root of the determinant of $\Pi(L)$ is on or inside the unit circle. Since we consider stationary variables, the two concepts are equivalent in our framework.

In standard situations, there is a one-to-one mapping between the u_t and ς_t and thus examining the fundamentalness of u_t provides information about the fundamentalness of ς_t. When the mapping is not one-to-one but the relationship between u_t and ς_t has a particular structure, it may be possible to find conditions insuring that when u_t is fundamental for x_t, ς_t is fundamental for χ_t, see e.g. Forni et al. (2009). In all other situations, many of which are of interest, knowing the properties of u_t for x_t may tells us little about the properties of the primitive shocks ς_t for χ_t.

Note that, although $\varsigma_{x,t}$ are linear combination of ς_t, they may still be economically interesting. An aggregate TFP shock may be meaningful, even if the sectoral TFP shocks drive the economy, as long as several sectoral TFP disturbances produce similar dynamics for the variables of the SVAR. On the other hand, it is not generally true that a fundamental shock is necessarily structurally interpretable (this occurs, for example, when the wrong D matrix is used to recover $\varsigma_{x,t}$ from a fundamental u_t).

3.1 Standard approaches to detect non-fundamentalness

Checking whether a Gaussian VAR is fundamental or not is complicated because the likelihood function or the spectral density can not distinguish between a fundamental and a non-fundamental representations. Earlier work by Lippi and Reichlin (1993, 1994) informally compared the dynamics produced by fundamental and selected non-fundamental representations. Giannone and Reichlin (2006) proposed to use Granger causality tests. The procedure works as follows. Suppose we augment x_t with a vector of variables y_t

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \Pi(L) & 0 \\ B(L) & C(L) \end{bmatrix} \begin{bmatrix} u_t \\ v_t \end{bmatrix} \tag{3.3}$$

where v_t are specific to y_t and orthogonal to u_t. Assume that all the roots of the determinant of $B(L)$ and $C(L)$ are outside the unit circle. If (3.2) is fundamental, $u_t = \Pi(L)^{-1}x_t$, and

$$y_t = B(L)\Pi(L)^{-1}x_t + C(L)v_t \tag{3.4}$$

where $B(L)\Pi(L)^{-1}$ is a one-sided in the non-negative powers of L. Thus, under fundamentalness, y_t is a function of current and past values of x_t, but x_t does not depend on y_t. Hence, to detect non-fundamentalness one can check whether x_t is predicted by lags of y_t.

While such an approach is useful to examine whether there are variables omitted from the empirical model, it is not clear whether it can reliably detect non-fundamentalness when shock aggregation is present. The reason is that cross-sectional aggregation is not innocuous. For example, Chang and Hong (2006) show that aggregate and sectoral technology shocks behave quite differently. The next example shows that aggregation may lead to spurious conclusions when using Granger causality to test for fundamentalness in small scale SVARs.

Example 4. Suppose the DGP is given by the following trivariate process:

$$\chi_{1t} = \varsigma_{1t} + b_1\varsigma_{1t-1} + a\varsigma_{2t} + a\varsigma_{3t} \tag{3.5}$$

$$\chi_{2t} = a\varsigma_{1t} + \varsigma_{2t} + b_2\varsigma_{2t-1} + a\varsigma_{3t} + \varsigma_{4t} \tag{3.6}$$

$$\chi_{3t} = a\varsigma_{1t} + a\varsigma_{2t} + \varsigma_{3t} + b_3\varsigma_{3t-1} - \varsigma_{4t} \tag{3.7}$$

where $\varsigma_t = [\varsigma_{1t}, \varsigma_{2t}, \varsigma_{3t}, \varsigma_{4t}]' \sim iid(0, \text{diag}(\Sigma_\varsigma))$ and $a \leq 1$.

Suppose an econometrician sets up a bivariate empirical model with $x_{1t} = \chi_{1t}$ and $x_{2t} = 0.5(\chi_{2t} + \chi_{3t})$. Thus, the second variable is an aggregated version of the last two variables of the DGP. The process generating x_t is

$$x_t = \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{bmatrix} 1 + b_1L & a & a \\ a & 0.5((a+1) + b_2L) & 0.5((a+1) + b_3L) \end{bmatrix} \begin{pmatrix} \varsigma_{1t} \\ \varsigma_{2t} \\ \varsigma_{3t} \end{pmatrix} \tag{3.8}$$

Because with two endogenous variables one can recover at most two shocks, the econometrician implicitly estimates:

$$x_t = \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{bmatrix} 1 + b_1L & a \\ a & 1 + cL \end{bmatrix} \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \tag{3.9}$$

where $\sigma^2_{u1} = \sigma^2_{\varsigma 1}$. Letting $\rho_0 + \rho_1 L \equiv [0.5(a+1)\ 0.5(a+1)] + [0.5b_2\ 0.5b_3]L$, and $\hat{\Sigma}_\varsigma = diag\{\sigma^2_{\varsigma 2}, \sigma^2_{\varsigma 3}\}$, c and σ^2_{u2} are obtained from:

$$\mathrm{E}(x_{2t}x'_{2t}) \equiv \gamma(0) = \rho_0\hat{\Sigma}_\varsigma\rho'_0 + \rho_1\hat{\Sigma}_\varsigma\rho'_1 = (1 + c^2)\sigma^2_{u2} \tag{3.10}$$

$$\mathrm{E}(x_{2t}x'_{2t-1}) \equiv \gamma(1) = \rho_1\hat{\Sigma}_\varsigma\rho'_0 = c\sigma^2_{u2} \tag{3.11}$$

These two conditions can be combined to obtain the quadratic equation:

$$c^2\gamma(1) - c\gamma(0) + \gamma(1) = 0 \tag{3.12}$$

Given $\gamma(0), \gamma(1)$ (3.12) can be used to compute the solution for c and then $\sigma^2_{u2} = c^{-1}\gamma(1)$.

Since u_t in (3.9) is a white noise, it is unpredictable using u_{t-s} (or x_{t-s}), $s > 0$. However, it can be predicted using ς_{t-s}, even when u_t is fundamental. In fact, letting c^* be the fundamental solution of (3.12) and using (3.8) and (3.9) have:

$$\begin{aligned} u_{2t} &= (1 + c^*L)^{-1}[\rho_0\hat{\varsigma}_t + \rho_1\hat{\varsigma}_{t-1}] \\ &= \rho_0\hat{\varsigma}_t + c^*\rho_0\hat{\varsigma}_{t-1} + (c^*)^2\rho_0\hat{\varsigma}_{t-2} + (c^*)^3\rho_0\hat{\varsigma}_{t-3} + \cdots \\ &+ \rho_1\hat{\varsigma}_{t-1} + c^*\rho_1\hat{\varsigma}_{t-2} + (c^*)^2\rho_1\hat{\varsigma}_{t-3} + (c^*)^3\rho_1\hat{\varsigma}_{t-4} + \cdots \end{aligned} \tag{3.13}$$

where $\hat{\varsigma} = [\varsigma_{2t}, \varsigma_{3t}]'$. Since χ_{2t-s} and χ_{3t-s} carry information about ς_{t-s}, lags of $y_t = [\chi_{2t}, \chi_{3t}]$ predict u_t, and thus x_t. Notice that in terms of equation (3.3), ς_{4t} plays the role of v_t. $\square$.

To gain intuition why predictability tests give spurious results notice that (3.13) implies $(1 + c^*L)u_{2t} = \rho_0\hat{\varsigma}_t + \rho_1\hat{\varsigma}_{t-1}$. Thus, under aggregation, estimated SVAR shocks are linear functions of current and **past** primitive structural shocks, making them predictable using any variable which has information about the lags of the primitive structural shocks. This occurs even if the VAR is correctly specified (i.e. there are sufficient lags to recover u_t as in (3.9)). In standard SVARs with no aggregation, the condition corresponding to (3.13) is $u_t = \rho\varsigma_t$. Thus, absent misspecification, lags of y_t will not predict u_t.

Granger causality tests have been used by many as a tool to detect misspecification in small scale VARs. For example, if a serially correlated variable is omitted from the VAR, the u_t the econometrician recovers are serially correlated and thus predictable using any variable correlated with the omitted one, see e.g. Canova et al. (2010). When they are applied to systems like those in example 4, causality tests detect misspecification but for the wrong reason. The VAR system is fundamental, the u_t derived from (3.9) are white noise, but Granger causality tests reject the predictability null because aggregation has created a particular correlation structure in SVAR shocks.

Example 4 also clearly highlights that the concepts of predictable, fundamental, and structural shocks are distinct. The u_t's in (3.9) are predictable, regardless of whether they are fundamental or not. In addition, $u_t = \varsigma_{x,t}$ are structural, in the sense that the responses of x_{1t} to u_t and to ς_{it}, $i = 1, 2, 3$, are similar, even u_t are predictable. Finally, u_t may be non-fundamental (if c, the non-fundamental solution of (3.12) is used in (3.13)), even if they are structural.

A similar outcome obtains if the empirical model contains, e.g., an estimated proxy for an observable variable or residuals computed from the elements of χ_t. Suppose $(x_{1t} = \chi_{1t}, x_{2t} = \chi_{1t} - \gamma_1\chi_{2t} - \gamma_2\chi_{3t})'$, and γ_1, γ_2 are (estimated) parameters. For example, x_{2t} are Solow residuals and γ_1, γ_2 are the labor and the capital shares. The process generating x_t is:

$$x_t = \begin{bmatrix} 1 + b_1 L & a & a & 0 \\ (1 - \gamma a - (1-\gamma)a) - b_1 L & (a - \gamma - a(1-\gamma)) - b_2 L & (a - \gamma a - (1-\gamma)) - b_3 L) & -\gamma_1 + \gamma_2 \end{bmatrix} \begin{pmatrix} \varsigma_{1t} \\ \varsigma_{2t} \\ \varsigma_{3t} \\ \varsigma_{4t} \end{pmatrix}$$

As before, the econometrician estimates (3.9). Also in this situation, u_t is unpredictable using u_{t-s} or x_{t-s}. However, lags of any y_t constructed as noisy linear transformation of $[\chi_{2t}, \chi_{3t}]$ predict u_t, even when it is fundamental for x_t.

In sum, the existence of variables that Granger cause x_t may have nothing to do with fundamentalness. What is crucial to create spurious results is that SVAR shocks linearly aggregate the information contained in current and past primitive structural shocks.

Although to some readers example 4 may look special, it is not. We next formally show that predictability obtains, in general, under linear cross-sectional aggregation. This together with the fact that small scale SVARs are generally used in business cycle analysis, even when the DGP may feature a large number of primitive structural shocks, should convince skeptical readers of the relevance of example 4. Proposition 1 shows that the class of moving average models is closed with respect to linear transformations and Proposition 2 that aggregated moving average models are predictable.

Proposition 1. Let χ_{1t} be a zero-mean MA(q_1) process:

$$\chi_{1t} = \varsigma_{1t} + \Phi_1 \varsigma_{1t-1} + \Phi_2 \varsigma_{1t-2} + \cdots + \Phi_{q_1} \varsigma_{1t-q_1} \equiv \Phi(L)\varsigma_{1t} \tag{3.14}$$

with $\mathrm{E}(\varsigma_{1t}\varsigma_{1t-j}) = \sigma_1^2$ if $j = 0$ and 0 otherwise,and let χ_{2t} be a zero-mean MA(q_2) process:

$$\chi_{2t} = \varsigma_{2t} + \Psi_1 \varsigma_{2t-1} + \Psi_2 \varsigma_{2t-2} + \cdots + \Psi_{q_2} \varsigma_{2t-q_2} \equiv \Psi(L)\varsigma_{2t} \tag{3.15}$$

with $\mathrm{E}(\varsigma_{2t}\varsigma_{2t-j}) = \sigma_2^2$ if $j = 0$ and 0 otherwise. Assume that χ_{1t} and χ_{2t} are independent at all leads and lags. Then

$$x_t = \chi_{1t} + \gamma\chi_{2t} = u_t + \Pi_1 u_{t-1} + \Pi_2 u_{t-2} + \cdots + \Pi_q u_{t-q} \equiv \Pi(L)u_t \tag{3.16}$$

where $q = \max\{q_1, q_2\}$, γ is a vector of constants, and u_t is a white noise process.

Proof: The proof follows from Hamilton (1994), page 106.□

Proposition 2. Let x_t be an m-dimensional process obtained as in Proposition 1. Then ς_{1t-s} and ς_{2t-s}, $s \geq 1$ Granger cause x_t.

Proof: It is enough to show that

$$\mathbb{P}\big[x_t | x_{t-1}, x_{t-2}, \cdots, \varsigma_{1t-1}, \varsigma_{1t-2}, \cdots, \varsigma_{2t-1}, \varsigma_{2t-2}, \cdots\big] \neq \mathbb{P}\big[x_t | x_{t-1}, x_{t-2}, \cdots\big]$$

when the model is fundamental, where $\mathbb{P}$ is the linear projection operator. Here $\mathcal{H}_t^x = \mathcal{H}_t^u$. Hence, it suffices to show that u_t is Granger caused by lagged values of ς_{1t} and ς_{2t}. That is

$$\mathbb{P}\big[u_t | u_{t-1}, u_{t-2}, \cdots, \varsigma_{1t-1}, \varsigma_{1t-2}, \cdots, \varsigma_{2t-1}, \varsigma_{2t-2}, \cdots\big] \neq \mathbb{P}\big[u_t | u_{t-1}, u_{t-2}, \cdots\big]$$

From Proposition 1, we have that $\Pi(L)u_t = \Phi(L)\varsigma_{1t} + \Psi(L)\varsigma_{2t}$, and therefore $u_t = \Pi(L)^{-1}\Phi(L)\varsigma_{1t} + \Pi(L)^{-1}\Psi(L)\varsigma_{2t}$, where $\Pi(L)^{-1}$ exists since the model is fundamental. Hence, $\Pi(L)^{-1}\Phi(L)$

and $\Pi(L)^{-1}\Psi(L)$ are one-sided polynomial in the non-negative powers of L and

$$\mathbb{P}[u_t|u_{t-1}, u_{t-2}, \cdots, \varsigma_{1t-1}, \varsigma_{1t-2}, \cdots, \varsigma_{2t-1}, \varsigma_{2t-2}, \cdots] = \mathbb{P}[u_t|\varsigma_{1t-1}, \varsigma_{1t-2}, \cdots, \varsigma_{2t-1}, \varsigma_{2t-2}, \cdots] \neq 0$$

where the equality follows from u_t being a white noise process. □

Thus, although u_t in (3.16) is unpredictable given own lagged values, it can be predicted using lagged values of ς_{1t} and ς_{2t} because the information contained in the histories of ς_{1t} and ς_{2t} is not optimally aggregated into u_t.

While the analysis is so far concerned with the fundamentalness of the vector u_t, it is common in the VAR literature to focus attention on just one shock. The next example shows when one can recover a shock from current and past values of the observables, even when the system is non-fundamental.

Example 5. Consider the following systems

$$\begin{aligned} x_{1,t} &= u_{1t} \\ x_{2,t} &= u_{1t} + u_{2t} - 3u_{2t-1} \end{aligned} \tag{3.17}$$

$$\begin{aligned} x_{1,t} &= u_{1t} - 2u_{2t-1} \\ x_{2,t} &= u_{1t-1} + u_{2t-1} \end{aligned} \tag{3.18}$$

Both systems are non-fundamental - the determinants of the MA matrix are $1 - 3z$, and $z(1 - 2z)$ respectively, and they both vanish for $z < 1$. Thus, it is impossible to recover $u_t = (u_{1t}, u_{2t})$ from current and lagged $x_t = (x_{1,t}, x_{2,t})'$. However, while in the first system u_{1t} can be obtained from $x_{1,t}$, in the second system no individual shock can not be obtained from linear combinations of current and past x_t's. □

A necessary condition for a SVAR shock to be an innovation is that it is orthogonal to the past values of the observables. FG suggest that a shock derived as in the first system of example 5 is fundamental if it is unpredictable using (orthogonal to the) lags of the principal components obtained from variables belonging to the econometrician's information set.

Three important points need to be made about such an approach. First, fundamentalness is a property of a system not of a single shock. Thus, orthogonality tests are, in general, insufficient to assess fundamentalness. Second, as it is clear from example 5, when one shock can be recovered, it is not the shock that creates non-fundamentalness in the first place.

Finally, an orthogonality test has the same shortcomings as a Granger causality test. It will reject the null of unpredictability of a SVAR shock using disaggregated variables or factors providing noisy information about them, when the SVAR shock is a linear combinations of primitive disturbances, for exactly the same reasons that Granger causality tests fail.

3.2 An alternative approach

In this section we propose an alternative testing approach that we expect to have better properties in the situations of interest in this paper. To see what the procedure involves suppose we still augment (3.2) with a vector of additional variables $y_t = B(L)u_t + C(L)v_t$. If (3.2) is fundamental, u_t can be obtained as from current and past values of x_t

$$u_t = x_t - \sum_{j=1}^{r} \omega_j x_{t-j} \tag{3.19}$$

where $\omega(L) = \Pi(L)^{-1}$ and r is generally finite. Thus, under fundamentalness y_t only depends on current and past values of u_t. If instead (3.2) is non-fundamental, u_t can not be recovered from the current and past values of the x_t. A VAR econometrician can only recover $u_t^* = x_t - \sum_{j=1}^{r} \omega_j^* x_{t-j}$, where $\omega(L)^* = \Pi(L)^{-1}\theta(L)^{-1}$, which is related to u_t via

$$u_t^* = \theta(L)u_t \tag{3.20}$$

where $\theta(L)$ is a Blaschke matrix [2]. Thus, the relationship between y_t and the shocks recovered by the econometrician is $y_t = B(L)\theta(L)^{-1}\theta(L)u_t + C(L)v_t \equiv B(L)^* u_t^* + C(L)v_t$. Since $B(L)^* \equiv B(L)\theta(L)^{-1}$ is generally a two-sided polynomial, y_t depends on current, past and *future* values of u_t^*. This proves the following proposition.

Proposition 3. The system (3.2) is fundamental if $u_{t+j}^*,\ j \geq 1$ fails to predict y_t.

Example 6. To illustrate proposition 3, let $x_t = (1-2.0L)u_t$, then:

$$x_t = (1-2.0L)\frac{(1-0.5L)}{(1-2.0L)}\frac{(1-2.0L)}{(1-0.5L)}u_t \equiv (1-0.5L)u_t^* \tag{3.21}$$

where $u_t^* = \frac{(1-2.0L)}{(1-0.5L)}u_t$. Let $y_t = (1-0.5L)u_t + (1-0.6L)v_t$. Then

$$\begin{aligned} y_t &= (1-0.5L)\frac{(1-0.5L)}{(1-2.0L)}u_t^* + (1-0.6L)v_t \\ &= \sum_{j=0}^{\infty}(1/2)^j((1-0.5L)^2 u_{t+j}^*) + (1-0.6L)v_{t-j} \end{aligned} \tag{3.22}$$

Two points about our testing procedure need to be stressed. First, Sims (1972) has shown that x_t is exogenous with respect to y_t if future values of x_t do not help to explain y_t. Similarly here, a VAR system is fundamental if future values of x_t (u_t) do not help to predict the variables y_t, excluded from the empirical model. Thus, although the null tested here and with Granger causality is the same, aggregation/non-observability problems may make the testing results different. Second, our approach is likely to have better size properties, when SVAR shocks are linear functions of lags of primitive shocks, because y_t generally contains more information than x_t - under fundamentalness, future values of u_t will not predict y_t. Note also that our test is sufficiently general to detect non-fundamentalness due to structural causes, omitted variables, or the use of proxy indicators.

[2] Blaschke matrices are complex-valued filters. The main property of Blaschke matrices is that they take orthonormal white noises into orthonormal white noises. See Lippi and Reichlin (1994) for more details.

4 Some Monte Carlo evidence

To evaluate the small sample properties of traditional predictability tests and of our new procedure, we carry out a simulation study using a version of the model of Leeper et al. (2013), with two sources of tax disturbances. The representative household maximizes:

$$\mathrm{E}_0 \sum_{t=0}^{\infty} \beta^t \log(C_t) \tag{4.1}$$

subject to

$$C_t + (1-\tau_{t,k})K_t + T_t \leq (1-\tau_{t,y})A_t K_{t-1}^{\alpha} = (1-\tau_{t,y})Y_t \tag{4.2}$$

where C_t, K_t, Y_t, T_t, $\tau_{t,k}$ and $\tau_{t,y}$ denote time-t consumption, capital, output, lump-sum transfers, investment tax and income tax rates, respectively; A_t is a technology disturbance and E_t is the conditional expectation operator. To keep the setup tractable, we assume full capital depreciation. The government sets tax rates randomly and adjusts transfers to satisfy $T_t = \tau_{t,y}Y_t + \tau_{t,k}K_t$. The Euler equation and the resource constraints are:

$$\frac{1}{C_t} = \alpha\beta\mathrm{E}_t\Big[\frac{(1-\tau_{t+1,y})}{(1-\tau_{t,k})}\frac{1}{C_{t+1}}\frac{A_{t+1}K_t^{\alpha}}{K_t}\Big] \tag{4.3}$$

$$C_t + K_t = A_t K_{t-1}^{\alpha} \tag{4.4}$$

Log linearizing, combining (4.3) and (4.4), we have

$$\hat{K}_t = \alpha\hat{K}_{t-1} + \sum_{i=0}^{\infty}\theta^i\mathrm{E}_t\hat{A}_{t+i+1} - \kappa_k\sum_{i=0}^{\infty}\theta^i\mathrm{E}_t\hat{\tau}_{t+i,k} - \kappa_y\sum_{i=0}^{\infty}\theta^i\mathrm{E}_t\hat{\tau}_{t+i+1,y} \tag{4.5}$$

where $\kappa_k = \frac{\tau_k(1-\theta)}{(1-\tau_k)}$, $\kappa_y = \frac{\tau_y(1-\theta)}{(1-\tau_y)}$, $\theta = \alpha\beta\frac{1-\tau_y}{1-\tau_k}$, $\hat{K}_t \equiv \log(K_t) - \log(K)$, $\hat{A}_t \equiv \log(A_t) - \log(A)$, $\hat{\tau}_{t,k} \equiv \log(\tau_{t,k}) - \log(\tau_k)$, $\hat{\tau}_{t,y} \equiv \log(\tau_{t,y}) - \log(\tau_y)$ and lower case letters denote percentage deviations from steady states.

We posit that technology and investment tax shocks are iid: $\hat{A}_t = \varsigma_{t,A}, \hat{\tau}_{t,k} = \varsigma_{t,k}$; and that the income tax shock is a MA(1) process: $\hat{\tau}_{t,y} = \varsigma_{t,y} + b\varsigma_{t-1,y}$. Then (4.5) is:

$$\hat{K}_t = \alpha \hat{K}_{t-1} + \varsigma_{t,a} - \kappa_k \varsigma_{t,k} - \kappa_y b \varsigma_{t,y} \tag{4.6}$$

We assume that an econometrician observes $\hat{K}_t$ and an aggregate tax variable:

$$\hat{\tau}_t = \omega \hat{\tau}_{t,y} + \hat{\tau}_{t,k} = \varsigma_{t,k} + \omega(\varsigma_{t,y} + b\varsigma_{t-1,y}) \tag{4.7}$$

where ω controls the relative weight of income taxes in the aggregate. Alternatively, one can assume that investment and income tax revenues are both observables, but an econometrician works with a weighted sum of them. If $(\hat{K}_t, \hat{\tau}_t)$ are the variables the econometrician uses in the VAR, our design covers both the cases of aggregation and of a relevant latent variable. In fact, the DGP for the observables is:

$$\begin{bmatrix} (1-\alpha L)\hat{K}_t \\ \hat{\tau}_t \end{bmatrix} = \begin{bmatrix} 1 & -\kappa_k & -\kappa_y b \\ 0 & 1 & \omega(1+bL) \end{bmatrix} \begin{bmatrix} \varsigma_{t,a} \\ \varsigma_{t,k} \\ \varsigma_{t,y} \end{bmatrix} \equiv \Gamma_x(L) C_x \varsigma_t \tag{4.8}$$

while the process recoverable by the econometrician is

$$\begin{bmatrix} (1-\alpha L)\hat{K}_t \\ \hat{\tau}_t \end{bmatrix} = \begin{bmatrix} 1 & \rho \\ 0 & 1+cL \end{bmatrix} \begin{bmatrix} u_{t,1} \\ u_{t,2} \end{bmatrix} \equiv \Pi(L) u_t \tag{4.9}$$

where $\sigma_1^2 = \sigma_a^2$ while c, σ_2^2, ρ are obtained from:

$$\begin{aligned} c^2 - c((1+b^2)/b + \sigma_k^2/(\omega^2 b \sigma_y^2)) + 1 &= 0 && (4.10) \\ \sigma_2^2 &= b\omega^2 \sigma_y^2 / c && (4.11) \\ \rho &= -\sqrt{(\omega^2 \kappa_y^2 b^2 \sigma_y^2 + \kappa_k^2 \sigma_k^2)/\sigma_2^2} && (4.12) \end{aligned}$$

By comparing (4.9) and (4.8), one can see that the aggregate tax shock $u_{t,2}$ will produce the same qualitative dynamic response in $\hat{K}_t$ as the investment and the income tax shocks but the scale of the effect will be altered. Depending on the size of ω, the aggregate shock will looks more like the income or the investment tax shock. For the exercises we present, we let $\varsigma_{t,a}, \varsigma_{t,k}, \varsigma_{t,y} \sim iid\ N(0,1)$; set $\alpha = 0.36$, $\beta = 0.99$, $\tau_y = 0.25$, $\tau_k = 0.1, \omega = 1$ and vary b so that $c \in (0.1, 0.8)$ (fundamentalness region) or $c \in (2, 9)$ (non-fundamentalness region).

To perform the tests, we need additional data not used in the empirical model (4.9). We assume that an econometrician observes a panel of 30 time series generated by:

$$(1 - 0.9L)y_{i,t} = \varsigma_{t,a} + \gamma_i \varsigma_{t,y} + (1 - \gamma_i)\varsigma_{t,k} + \xi_{i,t},\ i = 1, \cdots, 30 \tag{4.13}$$

where $\xi_{i,t} \sim\ iid\ N(0, \sigma_\xi^2)$, and γ_i is Bernoulli, taking value 1 with probability 0.5.

The properties of our procedure, denoted by CH, are examined with the regression:

$$f_t = \sum_{j=1}^{p_1} \phi_j f_{t-j} + \sum_{j=0}^{p_2} \psi_{-j} u_{t-j} + \sum_{j=1}^{q} \psi_j u_{t+j} + e_t \tag{4.14}$$

where f_t is a $s \times 1$ vector of principal components of (4.13) and u_t is estimated using

$$x_t = \sum_{j=1}^{r} \rho_j x_{t-j} + u_t \tag{4.15}$$

where $x_t = (\hat{\tau}_t, \hat{K}_t)'$. The null is $\mathbb{H}_0^{CH} : R\Psi = 0$, where $\Psi = \text{Vec}[\psi_1, \psi_2, \cdots, \psi_q]$, R is a matrix of zeros and ones. We report the results for $p_1 = 4, p_2 = 0, q = 2, r = 4$.

To examine the properties of Granger causality tests, denoted by GC, we employ

$$x_t = \sum_{j=0}^{p_1} \phi_j x_{t-j} + \sum_{j=1}^{p_2} \varphi_j f_{t-j} + e_t \tag{4.16}$$

where again $x_t = (\hat{\tau}_t, \hat{K}_t)'$. The null is $\mathbb{H}_0^{GC} : R\Phi = 0$ where $\Phi = \text{Vec}[\varphi_1, \varphi_2, \cdots, \varphi_{p_2}]$ and R is a matrix of zeros and ones. We report results for $p_1 = 4, p_2 = 2$.

To perform an orthogonality test, denoted by OR, we first estimate (4.15) with $r = 4$. The tax shock, $u_{t,\tau}$, is identified as the only one affecting $\hat{\tau}_t$. Then, in the regression

$$u_{t,\tau} = \sum_{j=1}^{p_2} \lambda_j f_{t-j} + e_t \tag{4.17}$$

the orthogonality null is $\mathbb{H}_0^{OR} : R\Lambda = 0$ where $\Lambda = \text{Vec}[\lambda_1, \lambda_2, \cdots, \lambda_q]$ and R is a matrix of zeros and ones. We report results for $p_2 = 2$.

To maintain comparability, all null hypotheses are tested using an F-test, setting $s = 3$ and $\sigma_\xi^2 = 1$ and no correction for generated regressors in (4.14) and (4.17). The appendix present results for the CH test when other values of p_2, σ_ξ^2, s, and q are used. We set $T = 200$, which is the length of the time series used in section 5, and $T = 2000$.

To better understand the properties of the tests, we also run an experiment with no aggregation problems. Here $\tau_{k,t} = 0, \forall t$, so that the DGP for capital and taxes is

$$\begin{bmatrix} (1-\alpha L)\hat{K}_t \\ \hat{\tau}_t \end{bmatrix} = \begin{bmatrix} 1 & -\kappa_y b \\ 0 & (1+bL) \end{bmatrix} \begin{bmatrix} \varsigma_{t,a} \\ \varsigma_{t,y} \end{bmatrix} \tag{4.18}$$

and the process for the additional data is

$$(1-0.9L)y_{i,t} = \varsigma_{t,a} + \gamma_i \varsigma_{t,y} + \xi_{i,t}, \; i = 1, \cdots, n \tag{4.19}$$

The percentage of rejections of the null in 1000 replications when the model is fundamental are in tables 1 and 2. Our procedure is undersized (it rejects less than expected from the nominal size) but its performance of independent of the nominal confidence level and the sample size. Granger causality and orthogonality tests are prone to spurious nonfundamentalness. This is clear when T=2000; in the smaller sample predictability due to aggregation is somewhat harder to detect.

Why are traditional predictability tests rejecting the null much more than one would

Table 1: Size of the tests: aggregation, T=200

	c	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	
	10%	1.5	1.3	0.5	0.6	1.4	1.7	1.2	1.7	
CH	5%	0.8	0.5	0.1	0.3	0.3	0.9	0.4	1.1	
	1%	0.1	0.1	0.1	0.2	0.1	0.2	0.1	0.1	
	10%	13.1	15.1	16.5	15.8	19.5	27.4	38.9	55.1	
GC	5%	7.5	8.2	9.5	9.2	11.2	15.5	26.7	42.2	
	1%	2.0	2.7	2.2	3.1	4.3	5.8	10.9	19.6	
	10%	5.2	5.7	5.3	6.2	6.5	6.2	8.5	13.2	
OR	5%	2.9	2.3	2.9	2.5	3.5	2.3	4.2	6.5	
	1%	0.1	0.5	0.6	0.2	0.2	0.7	0.7	1.6	

Notes: The table reports the percentage of rejections of the null hypothesis in 1000 replications when there is aggregation; CH is the test proposed in this paper; OR is the orthogonality test; GC is the Granger causality test. The length of the vector of principal components used in the testing equation is s=3.

Table 2: Size of the tests: aggregation, T=2000

	c	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	
	10%	0.1	0.4	0.2	0.1	0.2	0.1	1.9	9.5	
CH	5%	0.1	0.2	0.2	0.1	0.1	0.1	1.0	4.2	
	1%	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.8	
	10%	83.3	86.6	88.8	92.4	98.1	99.9	100	100	
GC	5%	75.3	75.1	76.3	83.9	96.1	99.8	100	100	
	1%	49.7	44.7	46.0	58.7	83.9	98.6	100	100	
	10%	34.0	30.1	29.4	30.2	41.7	52.1	81.0	99.1	
OR	5%	21.4	18.5	18.5	19.0	27.9	36.0	66.4	96.5	
	1%	7.4	6.8	7.3	6.4	9.0	13.0	34.9	81.8	

Notes: The table reports the percentage of rejections of the null hypothesis in 1000 replications when there is aggregation; CH is the test proposed in this paper; OR is the orthogonality test; GC is the Granger causality test. The length of the vector of principal components used in the testing equation is s=3.

Table 3: Size of the tests: no aggregation, T=200

	c	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
	10%	2.5	1.7	2.5	1.6	1.5	2.3	2.8	2.6
CH	5%	1.1	0.6	0.6	1.2	0.5	0.8	1.2	1.0
	1%	0.1	0.1	0.2	0.1	0.1	0.1	0.2	0.3
	10%	11.4	10.5	13.3	13.5	10.9	14.8	15.5	28.4
GC	5%	5.6	5.0	6.2	8.2	5.3	7.4	9.2	19.8
	1%	1.3	1.0	1.6	1.6	1.1	2.0	2.4	6.1
	10%	4.4	5.1	5.3	4.7	4.0	6.4	6.2	8.9
OR	5%	1.7	1.3	2.8	2.2	1.3	2.3	2.6	4.9
	1%	0.2	0.1	0.5	0.3	0.1	0.6	0.6	1.8

Notes: The table reports the percentage of rejections of the null hypothesis in 1000 replications when there is no aggregation; CH is the test proposed in this paper; OR is the orthogonality test; GC is the Granger causality test. The length of the vector of principal components used in the testing equation is s=3.

Table 4: Size of the tests: no aggregation, T=2000

	c	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
	10%	0.9	1.3	1.0	0.9	0.5	1.0	1.3	6.0
CH	5%	0.3	0.5	0.5	0.3	0.2	0.3	0.5	3.8
	1%	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1.7
	10%	13.2	13.3	15.6	14.8	18.2	26.7	53.8	99.8
GC	5%	7.4	8.0	8.7	9.0	11.1	16.3	41.2	99.3
	1%	1.6	1.9	2.5	3.1	3.0	5.2	19.4	95.3
	10%	3.9	5.2	5.6	4.8	4.2	6.8	8.7	20.9
OR	5%	1.3	3.2	1.7	1.8	1.7	3.2	4.8	17.8
	1%	0.3	0.7	0.4	0.4	0.2	0.4	1.2	6.0

Notes: The table reports the percentage of rejections of the null hypothesis in 1000 replications when there is no aggregation; CH is the test proposed in this paper; OR is the orthogonality test; GC is the Granger causality test. The length of the vector of principal components used in the testing equation is s=3.

Table 5: Power of the tests: aggregation, T=200

	c	2	3	4	5	6	7	8	9
	10%	99.9	100	100	100	100	100	100	100
CH	5%	99.5	100	100	100	100	100	100	100
	1%	98.7	100	100	100	100	100	100	100
	10%	100	100	100	100	100	100	100	100
GC	5%	100	100	100	100	100	100	100	100
	1%	100	100	100	100	100	100	100	100
	10%	100	100	100	100	100	100	100	100
OR	5%	100	100	100	100	100	100	100	100
	1%	100	100	100	100	100	100	100	100

Notes: The table reports the percentage of rejections of the null hypothesis in 1000 replications when there is aggregation; CH is the test proposed in this paper; OR is the orthogonality test; GC is the Granger causality test. The length of the vector of principal components used in the testing equation is s=3.

Table 6: Power of the tests: no aggregation, T=200

	c	2	3	4	5	6	7	8	9
	10%	100	100	100	100	100	100	100	100
CH	5%	100	100	100	100	100	100	100	100
	1%	100	100	100	100	100	100	100	100
	10%	100	100	100	100	100	100	100	100
GC	5%	100	100	100	100	100	100	100	100
	1%	100	100	100	100	100	100	100	100
	10%	100	100	100	100	100	100	100	100
OR	5%	100	100	100	100	100	100	100	100
	1%	100	100	100	100	100	100	100	100

Notes: The table reports the percentage of rejections of the null hypothesis in 1000 replications when there is no aggregation; CH is the test proposed in this paper; OR is the orthogonality test; GC is the Granger causality test. The length of the vector of principal components used in the testing equation is s=3.

expect from the nominal size? The answer is obtained recalling equation (3.13). u_t are linear combinations of current and past values of $\hat{A}_t, \hat{\tau}_{t,k}, \hat{\tau}_{t,y}$ while f_t are linear combinations of $\hat{A}_t, \hat{\tau}_{t,k}, \hat{\tau}_{t,y}$ and $\xi_{i,t}, i = 1, \ldots, 30$. Since $\hat{\tau}_{t,k}$ is serially correlated, lags of f_t may help to predict x_t even when lags of x_t are included, in particular, when the draws for γ_i are small.

It is known that Granger causality tests have poor size properties when x_t is persistent, see e.g. Ohanian (1988). Tables 3 and 4 disentangle aggregation from persistence problems: since they have been constructed absent aggregation, they report size distortions due to persistent data. It is clear that, when $b > 0.6$, the size of Granger causality tests is distorted. To properly run such tests, the lag length p_1 of the testing equation must be made function of the (unknown) persistence of the DGP. However, when $b > 0.8$, distortions are present even if $p_1 = 10$. The orthogonality test performs better because it preliminary filters x_t with a VAR. Thus, high serial correlation in x_t is less of a problem.

Comparing the size tables constructed with and without aggregation, one can see that the properties of the CH test do not depend on the presence of aggregation or the persistence of the DGP. On the other hand, aggregation make the properties of Granger causality and orthogonality tests significantly worse.

Tables 5 and 6 report the empirical power of the tests when T=200 with and without aggregation. All tests are similarly powerful to detect non-fundamentalness when it is present, regardless of the confidence level and the nature of the DGP. Although not reported for reasons of space, the power of the three tests is unchanged when T=2000.

The additional tables in the appendix indicate that the size properties of the CH test are insensitive to the selection of three nuisance parameters: the variance of the shocks to the additional data σ^2_ξ, the number of principal components used in the testing equation s, and the number of leads of the first stage residuals used in the testing equation q. On the other hand, the choice of p_2, the number of lags of the first stage residuals used in the testing equations, matters. This is true, in particular, when the persistence of the DGP increases and is due to the fact that with high persistence, r=4 is insufficient to whiten the first stage residual, and the presence of serial correlation in u_t makes its future values spuriously significant. To avoid this problem in practice, we recommend users to specify the testing equation with only leads of u_t. Alternatively, if lags of u_t are included, r should be large to insure that serial correlation in the first stage residuals is negligible.

5 Reconsidering a small scale SVAR

Standard business cycle theories assume that economic fluctuations are driven by surprises in current fundamentals, such as aggregate productivity or the monetary policy rule. Motivated by the idea that changes in expectations about future fundamentals may drive business fluctuations, Beaudry and Portier (2006) study the effect of news shocks on the real economy using a SVAR that contains stock prices and TFP.

Since models featuring news shocks have solutions displaying moving average components, empirical models with a finite number of lags may be unable to capture the underlying dynamics, making the SVARs considered in the literature prone to non-fundamentalness. In addition, Forni et al. (2014) provide a stylized Lucas tree model where perfectly predictable news to the dividend process may induce non-fundamentalness in a VAR system comprising the growth rate of stock prices and the growth rate of dividends. The solution of their model, when news come two periods in advance is:

$$\begin{bmatrix} \Delta d_t \\ \Delta p_t \end{bmatrix} = \begin{bmatrix} L^2 & 1 \\ \frac{\beta^2}{1-\beta} + \beta L & \frac{\beta}{1-\beta} \end{bmatrix} \begin{bmatrix} \varsigma_{1t} \\ \varsigma_{2t} \end{bmatrix} \equiv C(L)\varsigma_t \tag{5.1}$$

where d_t are dividends, p_t are stock prices, $0 < \beta < 1$ is the discount factor. Since $|C(L)|$ vanishes for $L = 1$ and $L = -\beta$, u_t is non-fundamental for $(\Delta d_t, \Delta p_t)$. Intuitively, this occurs because agents' information set, which includes current and past values of structural shocks, is not aligned with the econometrician's information set, which includes current and past values of the growth rate of dividends and stock prices. The fundamental and non-

Table 7: Testing fundamentalness: VAR with TFP growth and stock prices growth.

	PC=3	PC=4	PC=5	PC=6	PC=7	PC=8	PC=9	PC=10
sample 1960-2010								
CH	0.05	0.08	0.03	0.15	0.13	0.08	0.14	0.13
GC	0.02	0.00	0.04	0.01	0.00	0.00	0.00	0.00
Fernald data, sample 1960-2005								
GC(agg)	0.02	0.16	0.22	0.00	0.00	0.02	0.01	0.01
GC(dis)	0.17	0.52	0.54	0.11	0.09	0.17	0.25	0.34
Wang data, sample 1960-2009								
GC(agg)	0.05	0.02	0.11	0.03	0.00	0.00	0.00	0.00
GC(dis)	0.37	0.38	0.51	0.40	0.27	0.28	0.27	0.23

Notes: The table reports the p-value of the tests; CH is the test proposed in this paper; GC is the Granger causality test; the row GC(agg) reports the results of the test using aggregate data, the row GC(dis) the results of the test using disaggregated data; PC is the number of principal component in the auxiliary regression. In CH test the number of leads tested is two and the preliminary VAR has 4 lags. In GC test the lag length of the VAR is chosen with BIC and two lags of the principal components are used in the tests.

fundamental dynamics this model generates in response to news shocks are similar because the root generating non-fundamentalness ($L = -\beta$) is near unity, see also Beaudry et al. (2015). In general, the properties of the SVAR the econometrician considers depend on the process describing the information flows, on the variables observed by the econometrician and those included in the SVAR.

To reexamine the evidence we estimate a VAR with the growth rates of capacity adjusted TFP and of stock prices for the period 1960Q1 to 2010Q4, both of which are taken from Beaudry and Portier (2014) and we use the same principal components as in Forni et al. (2014). Table 7 reports the $p-$values of the tests, varying the number of principal components employed in the auxiliary regression, which enter in first difference in all the tests. In the CH test, the testing model has four lags of the PC and we are examining the predictive power of 2 leads of the VAR residuals. In the GC test the lag length of the VAR is chosen by BIC and two lags of the principal components are used in the tests.

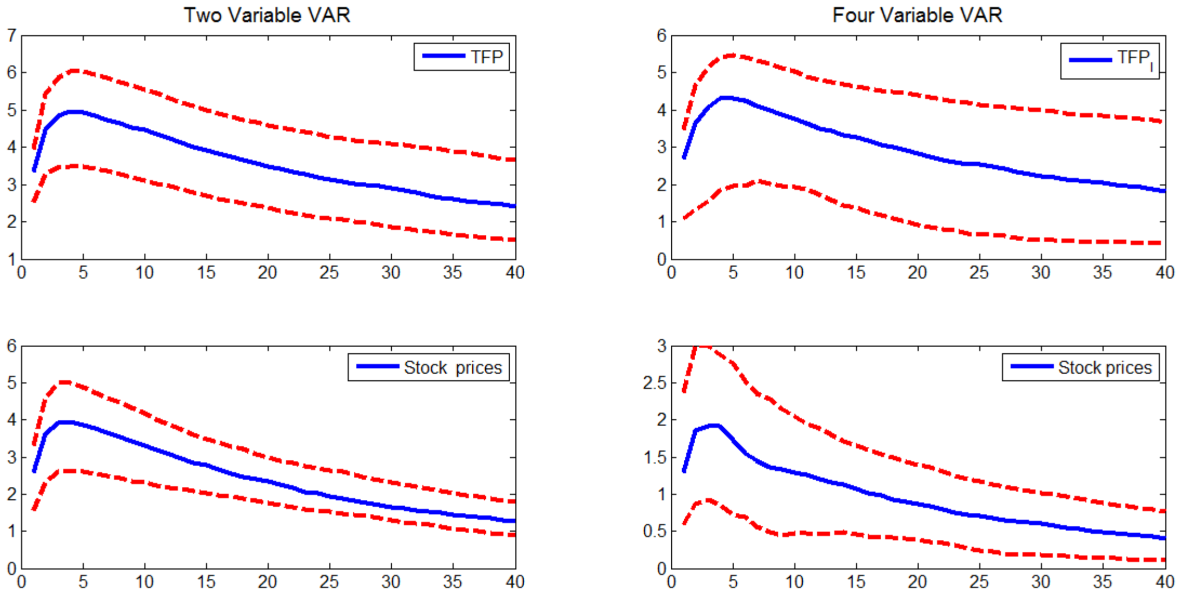
The CH test finds the system fundamental and, in general, the number of PC included in the testing equations does not matter. In contrast, a Granger causality test rejects the null of fundamentalness. Since the VAR includes TFP, which is a latent variable, and estimates are obtained from an aggregated production function, differences in the results could be due to aggregation and/or non-observability problems.

To verify this possibility we consider a VAR where in place of utilization adjusted *aggregated* TFP we consider two different utilization adjusted *sectoral* TFP measures. The first was constructed by John Fernald at the Federal Reserve Bank of San Francisco, and is obtained using the methodology of Fernald et al. (2013), which produces time series for private consumption TFP, private investment TFP, government consumption and investment TFP and ' net trade" TFP. The second panel of table 7 presents results obtained in a VAR which includes consumption TFP (obtained aggregating private and public consumption), investment TFP (obtained aggregating private and public investments) and net trade TFP, all in log growth rates, and the growth rate of stock prices. Because the data ends in 2005, the first row of the panel reports the p-values of a Granger causality test for the original bivariate system restricted to the 1960-2005 sample.

As an alternative, we use the utilization adjusted *industry* TFP data constructed by Christina Wang at the Federal Reserve Bank of Boston. We reaggregate industry TFPs into manufacturing, services and 'others' sectors, convert the data from annual to quarterly using a polynomial regression and use the growth rate of these three variables together with the growth rate of stock prices in the VAR. The third panel of table 7 presents results obtained with this VAR. Because the data ends in 2009, the first row of the panel reports the p-values of a Granger causality test for the original bivariate system restricted to the 1960-2009 sample.

Granger causality tests applied to the original bivariate system estimated over the two new samples still find the VAR non-fundamental. When the test is used in the VARs with sectoral/industry TFP measures, the null of non-fundamentalness is instead not rejected for all choices of vectors of principal components. Since this result holds when we enter the

Figure 1: Responses to technology news shocks



Note: The dotted regions report pointwise 68 % credible intervals; the solid line is the pointwise median response. The x-axis reports quarters, the y-axis the response of the level of the variable in deviation from the predictable path.

sectoral/industry TFP variables in level rather than growth rates, when we allow for a break in the TFP series, and when we use only two sectoral/industry TFP variables in the VAR, the conclusion is that a Granger causality test rejects the null in the original VAR because of aggregation problems. The diagnostic of this paper, being robust to aggregation problems, correctly identifies the original bivariate VAR as fundamental.

Clearly, if the DGP is a truely sectoral model, the shocks and the dynamics produced by both the bivariate and the four variable VAR systems are likely to be averages of the shocks and dynamics of the primitive economy, which surely includes more than two or four disturbances. The interesting question is whether the news shocks extracted in the two and four variable systems produce different TFP responses.

For illustration, figure 1 reports the responses of stock prices and of TFP to standardized technology news shocks in the original VAR and in the four varibale VAR with Fernald disaggregated TFP measures. For the four variable VAR we only present the responses of investment TFP since the responses of the other two TFP variables are insignificantly differ-

Figure 1: Responses to technology news shocks

Note: The dotted regions report pointwise 68 % credible intervals; the solid line is the pointwise median response. The x-axis reports quarters, the y-axis the response of the level of the variable in deviation from the predictable path.

and shocks. In this situation, SVAR shocks are linear transformations of current and past primitive structural shocks perturbing the economy. SVAR shocks might be fundamental or non-fundamental, depending on the details of the economy, the information set available to the econometrician, and the variables chosen in the empirical analysis. However, variables providing noisy information about the primitive structural shocks will Granger cause SVAR shocks, even when the SVAR is fundamental. A similar problem arises when SVAR variables proxy for latent variables. We conduct a simulation study illustrating that *spurious non-fundamentalness* may indeed occur when the SVAR used for the empirical analysis is of smaller scale than the DGP of the data.

We propose an alternative testing procedure which has the same power properties as existing diagnostics when non-fundamentalness is present, but does not face aggregation or non-observability problems when the system is fundamental. We also show that the ent from zero. It is clear that the conditional dynamics in the two systems are qualitatively similar and statistically indistinguishable. Nevertheless, median responses are smaller, uncertainty is more pervasive, and the hump in the TFP response muted in the larger system. Hence, cross sectional aggregation does not change much the dynamics but makes TFP responses artificially large and more precisely estimated. Researchers often construct models to quantitatively match the dynamics induced by shocks in small scale VARs. Figure 1 suggests that the size and the persistence of the structural shocks needed to produce the aggregate evidence are probably smaller than previously agreed upon.

6 Conclusions

Small scale SVAR models are often used in empirical business cycle analyses even though the economic model one thinks has generated the data has a larger number of variables and shocks. In this situation, SVAR shocks are linear transformations of current and past primitive structural shocks perturbing the economy. SVAR shocks might be fundamental or non-fundamental, depending on the details of the economy, the information set available to the econometrician, and the variables chosen in the empirical analysis. However, variables providing noisy information about the primitive structural shocks will Granger cause SVAR shocks, even when the SVAR is fundamental. A similar problem arises when SVAR variables proxy for latent variables. We conduct a simulation study illustrating that *spurious non-fundamentalness* may indeed occur when the SVAR used for the empirical analysis is of smaller scale than the DGP of the data.

We propose an alternative testing procedure which has the same power properties as existing diagnostics when non-fundamentalness is present, but does not face aggregation or non-observability problems when the system is fundamental. We also show that the procedure is robust to specification issues and to nuisance features. We demonstrate that a Granger causality diagnostic finds that a bivariate SVAR measuring the impact of news is non-fundamental, while our test finds it fundamental. The presence of an aggregated TFP measure in the SVAR explains the discrepancy. When sectoral TFP measures are used, a Granger causality diagnostic also finds the SVAR fundamental.

A few lessons can be learned from our paper. First, Granger causality tests may give misleading conclusions when testing for fundamentalness whenever aggregation or non-observability problems are present. Second, to derive reliable conclusions, one should have fundamentalness tests that are insensitive to specification and nuisance features. The test proposed in this paper satisfies both criteria; those present in the literature do not. Finally, if one is willing to assume that the DGP is a particular structural model, the procedure described Sims and Zha (2006) can be used to check if a particular VAR shock can be recovered from current and past values of the observables, therefore by-passing the need to check for fundamentalness. However, when the DGP is unknown, the structural model one employs misspecified, or the exact mapping from the DGP and the estimated SVAR hard to construct, procedures like ours can help researchers to understand whether small scale SVARs are good starting points to undertake informative business cycle analyses.

References

Lucia Alessi, Matteo Barigozzi, and Marco Capasso. Non-fundamentalness in structural econometric models: A review. *International Statistical Review*, 79(1):16–47, 2011.

Paul Beaudry and Franck Portier. Stock prices, news, and economic fluctuations. *The American economic review*, 96(4):1293–1307, 2006.

Paul Beaudry and Franck Portier. News-driven business cycles: insights and challenges. *Journal of Economic Literature*, 52(4):993–1074, 2014.

Paul Beaudry, Patrick Fève, Alain Guay, and Franck Portier. When is nonfundamentalness in vars a real problem? an application to news shocks. Technical report, National Bureau of Economic Research, 2015.

Fabio Canova. *Methods for applied macroeconomic research*, volume 13. Princeton University Press, 2007.

Fabio Canova, David Lopez-Salido, and Claudio Michelacci. The effects of technology shocks on hours and output: A robustness analysis. *Journal of Applied Econometrics*, 25(5):755–773, 2010.

Yongsung Chang and Jay H Hong. Do technological improvements in the manufacturing sector raise or lower employment? *The American economic review*, 96(1):352–368, 2006.

Jon Faust and Eric M Leeper. When do long-run identifying restrictions give reliable results? *Journal of Business & Economic Statistics*, 15(3):345–353, 1997.

John Fernald, Jonas Fisher, and Miles Kimball. Sector-specific technical change. 2013.

Jess Fernndez-Villaverde, Juan F. Rubio-Ramrez, Thomas J. Sargent, and Mark W. Watson. Abcs (and ds) of understanding vars. *American Economic Review*, 97(3):1021–1026, June 2007.

Mario Forni and Luca Gambetti. Sufficient information in structural vars. *Journal of Monetary Economics*, 66:124–136, 2014.

Mario Forni, Domenico Giannone, Marco Lippi, and Lucrezia Reichlin. Opening the black box: Structural factor models with large cross sections. *Econometric Theory*, 25(05): 1319–1347, 2009.

Mario Forni, Luca Gambetti, and Luca Sala. No news in business cycles. *The Economic Journal*, 2014.

Raffaella Giacomini. The relationship between dsge and var models. In *VAR Models in Macroeconomics–New Developments and Applications: Essays in Honor of Christopher A. Sims*, pages 1–25. Emerald Group Publishing Limited, 2013.

Domenico Giannone and Lucrezia Reichlin. Does information help recovering structural shocks from past observations? *Journal of the European Economic Association*, 4(2-3): 455–465, 2006.

Mehdi Hamidi Sahneh. Are the shocks obtained from svar fundamental? 2014.

James Douglas Hamilton. *Time series analysis*. Princeton university press Princeton, 1994.

Lars Peter Hansen and Robert J Hodrick. Forward exchange rates as optimal predictors of future spot rates: An econometric analysis. *journal of Political Economy*, 88(5):829–853, 1980.

Lars Peter Hansen and Thomas J Sargent. Formulating and estimating dynamic linear rational expectations models. *Journal of Economic Dynamics and control*, 2:7–46, 1980.

Lars Peter Hansen and Thomas J Sargent. Two difficulties in interpreting vector autoregressions. *Rational expectations econometrics*, pages 77–120, 1991.

Lutz Kilian and Helmut Lütkepohl. Structural vector autoregressive analysis, 2016.

Lutz Kilian and Daniel P Murphy. The role of inventories and speculative trading in the global market for crude oil. *Journal of Applied Econometrics*, 29(3):454–478, 2014.

Eric M Leeper, Todd B Walker, and Shu-Chun Susan Yang. Fiscal foresight and information flows. *Econometrica*, 81(3):1115–1145, 2013.

Keh-Shin Lii and Murray Rosenblatt. Deconvolution and estimation of transfer function phase and coefficients for nongaussian linear processes. *The Annals of Statistics*, pages 1195–1208, 1982.

Marco Lippi and Lucrezia Reichlin. The dynamic effects of aggregate demand and supply disturbances: reply. *The American Economic Review*, 83(0):653–658, 1993.

Marco Lippi and Lucrezia Reichlin. Var analysis, nonfundamental representations, blaschke matrices. *Journal of Econometrics*, 63(1):307–325, 1994.

Helmut Lütkepohl. *Introduction to multiple time series analysis*. Springer and Verlag, 1991.

Lee E Ohanian. The spurious effects of unit roots on vector autoregressions: A monte carlo study. *Journal of Econometrics*, 39(3):251–266, 1988.

Federico Ravenna. Vector autoregressions and reduced form representations of dsge models. *Journal of monetary economics*, 54(7):2048–2064, 2007.

Y. A. Rozanov. *Stationary random processes*. Holden-Day, 1967.

Christopher A Sims. Money, income, and causality. *The American economic review*, 62(4): 540–552, 1972.

Christopher A Sims and Tao Zha. Does monetary policy generate recessions? *Macroeconomic Dynamics*, 10(02):231–272, 2006.

Appendix

This appendix reports the size of the CH test when nuisance parameters are varied. We change the number of lags of first stage residuals in the auxiliary regression p_2; the variance of the error in the DGP for the additional variables, σ_ξ^2; the number of principal components used in the auxiliary regressions, s, the number of leads of the first stage residuals in the auxiliary regression q. Power tables are omitted, since they identical to those in the text.

Table A1: Size of the CH test, aggregation, varying p_2

	c	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
	10%	11.2	13.5	14.5	13.3	14.8	20.1	29.0	44.1
$p_2 = 4$	5%	2.5	2.3	2.5	2.2	2.9	4.6	6.1	11.9
	1%	1.6	1.9	1.2	1.6	2.2	4.1	6.2	12.3
	10%	10.5	13.2	12.1	12.5	14.1	19.3	27.0	40.8
$p_2 = 2$	5%	5.8	7.1	5.4	6.0	7.6	12.2	15.9	29.7
	1%	1.8	2.0	0.9	1.1	2.1	3.2	5.7	12.5

Notes: The table reports the percentage of rejections of the null hypothesis in 1000 replications when there is aggregation, T=200, and three principal components of the large dataset are considered; p_2 represents the number of lags in the testing equation (4.14).

Table A2: Size of the CH-test, aggregation, varying σ_ξ^2

	c	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
	10%	2.20	1.80	1.70	2.10	1.60	2.10	1.80	3.00
$\sigma_\xi^2 = 4$	5%	1.10	0.70	0.40	0.60	0.50	1.00	0.60	0.90
	1%	0.30	0.10	0.10	0.00	0.00	0.20	0.20	0.10
	10%	1.00	0.70	0.20	0.80	0.50	1.50	0.60	1.10
$\sigma_\xi^2 = 0.25$	5%	0.50	0.40	0.10	0.20	0.40	0.50	0.30	0.30
	1%	0.00	0.20	0.00	0.10	0.00	0.00	0.10	0.10

Notes: The table reports the percentage of rejections of the null hypothesis in 1000 replications when there is aggregation, T=200, and three principal components of the large dataset are considered; σ_ξ^2 is the variance of the idiosyncratic error in the DGP for additional data.

Table A3: Size of the CH test, aggregation, varying s

	c	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
	10%	1.10	1.10	0.30	0.60	0.80	1.00	1.00	1.70
$s = 2$	5%	0.50	0.50	0.00	0.30	0.40	0.50	0.10	0.60
	1%	0.10	0.10	0.00	0.10	0.00	0.10	0.00	0.10
	10%	1.70	1.80	0.70	1.80	1.40	1.90	1.40	2.50
$s = 4$	5%	0.80	0.70	0.10	0.60	0.50	0.60	0.50	1.10
	1%	0.20	0.10	0.00	0.10	0.00	0.20	0.10	0.10

Notes: The table reports the percentage of rejections of the null hypothesis in 1000 replications when there is aggregation, T=200, and three principal components of the large dataset are considered; s is the length of the vector of factors in the testing equation (4.14).

Table A4: Size of the CH test, aggregation, varying q

	c	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
	10%	1.80	3.10	2.40	1.90	2.00	2.60	1.60	3.70
$q = 1$	5%	0.70	1.40	0.80	0.30	0.70	1.50	0.70	2.10
	1%	0.00	0.10	0.00	0.00	0.40	0.10	0.30	0.50
	10%	1.20	0.80	0.50	0.70	0.90	1.20	0.60	1.80
$q = 2$	5%	0.40	0.20	0.20	0.30	0.30	0.50	0.30	0.80
	1%	0.00	0.00	0.00	0.00	0.00	0.10	0.00	0.10

Notes: The table reports the percentage of rejections of the null hypothesis in 1000 replications when there is aggregation, T=200, and three principal components of the large dataset are considered; q represents the number of leads in the testing equation (4.14).

Are the shocks obtained from SVAR fundamental?

Mehdi Hamidi Sahneh *†

Abstract

Non-fundamentalness arises when observed variables do not contain enough information to recover structural shocks. This paper provides new conditions under which the shocks recovered from a SVAR are fundamental. I prove that the Wold innovations are unpredictable if and only if the model is fundamental. I propose a test based on a generalized spectral density to check the unpredictability of the Wold innovations. I apply the test to study the dynamic effects of government spending on economic activity. I find that standard SVAR models commonly employed in the literature are non-fundamental. Moreover, I formally check if introducing a narrative variable that measures anticipation restores fundamentalness.

Keywords: Fiscal Policy; Fundamentalness; Identification; Invertibility; Vector Autoregressive.

JEL classification: C5, C32, E62.

---

*Departamento de Economia, Universidad Carlos III de Madrid, Getafe, 28903, Spain. Email address: *mhamidis@eco.uc3m.es*. The author is deeply indebted to Carlos Velasco for guidance and encouragement. I also benefited most from the comments of Fabio Canova, Juan Dolado, Jesus Gonzalo, Hernan Seoane, and seminar participants at Carlos III university, The 68th European Meeting of the Econometric Society (ESEM2014), 69th European Winter Meetings of the Econometric Society (EWM2014), The 8th Nordic Econometric Meeting (NEM2015), for comments and discussions. The research was supported by the Spanish Plan Nacional de I+D+I (ECO2012-31748) and (ECO2014-57007).

†Early versions of this paper has been presented at UC3M student workshop (September 2013), and has been circulated online. The current version is essentially the same with minor corrections.

1 Introduction

Structural Vector Autoregressive (SVAR) models have been used extensively for economic analysis. The underlying assumption of SVAR, known as fundamentalness, is that one is able to recover the structural shocks driving the process from linear combinations of observed present and past values of the process. Non-fundamentalness arises when observed variables do not contain enough information to recover the structural shocks and the impulse response functions. Once the representation is non-fundamental, all identification schemes, such as long-run or sign restrictions, fail to recover the true structural shocks. In this paper, I propose a test to empirically detect whether the shocks recovered from the estimation of a VAR are truly fundamental.

Although many economic models generate non-fundamental representations, little is known how to test if a model is non-fundamental. Permanent income models (Fernández-Villaverde et al., 2007), news shocks (Blanchard et al., 2013; Forni et al., 2014), and fiscal foresight (Leeper et al., 2013) are some examples that can generate equilibrium solutions with non-fundamental representation. For a comprehensive survey of this literature see Alessi et al. (2011).

The key contribution of this paper is to provide new conditions under which the shocks obtained from the estimates of the SVAR are truly fundamental. I prove that the Wold innovations from fitting a VAR to a non-fundamental model are not martingale difference and therefore predictable (in the mean), even if one includes an infinite past of the observable variables. Consequently, to test whether the model is fundamental, one must check if the Wold innovations are unpredictable.

There are some proposals to test for the unpredictability of the Wold innovations (see Hong (1999), Domínguez and Lobato (2003), Hong and Lee (2005), Escanciano and Velasco (2006), among others). To the best of my knowledge, none of these tests are applicable to the multivariate setting of this paper. Alternatively, it is possible to apply a sequence of univariate test to each series. However, using a multivariate procedure avoids the multiple testing problem and is more powerful, since it is possible that a single series is unpredictable, but the collection of several series is predictable. To test for the unpredictability of the Wold innovations, I extend Hong and Lee's (2005) test from univariate to multivariate setting. I show that the proposed test statistic has a convenient asymptotic standard normal distribution and diverges to infinity under the alternative hypothesis. The proposed test is simple to apply since it only needs reduced form VAR residuals as input. Therefore, my proposed test does not require any identification assumption or estimating non-fundamental models. I perform a Monte Carlo exercise, using a version of fiscal foresight model of Leeper et al. (2013) as DGP to examine the properties of the procedure. Simulations show that the test has good power against general alternatives. I also show that increasing the lag order of the estimated VAR does not solve the non-fundamentalness problem.

This paper is related to the literature that attempts to test if a Vector Moving Average (VMA) model is fundamental. Giannone and Reichlin (2006) prove that if a model is fundamental, then extra information should not Granger cause the variables included in the model. Similarly, Forni and Gambetti (2014) exploit the factors of a large system to propose necessary and sufficient conditions under which a VAR contains sufficient information to estimate the structural shocks, which under some assumptions could be applied to detect fundamentalness. However, these procedures are based on the untestable assumption that the extra information -such as sectoral data or factors of a large data set- that one uses to test for fundamentalness is itself fundamental.

From a methodological point of view, my proposal is similar to Chen et al. (2012). By converting testing for fundamentalness to testing for serial independence of the Wold innovations, these authors proposed a test for fundamental VMA representation. However, their test critically depends on the *iid* assumption of the true unobserved errors, which is often rejected in macroeconomic and financial time series. Failure to accommodate these features will lead to rejection of the null of fundamentalness by mistake. In contrast, my proposal is robust to the failure of the *iid* assumption.

To illustrate the application of the proposed test, I focus on the dynamic effects of government spending shocks on economic activity in the United States in the post-war period. I find that the baseline VAR models normally considered in the empirical literature to identify these effects are non-fundamental, and therefore, the impulse responses and variance decompositions from SVAR approach appears not to be reliable. In case of rejection of the null of fundamentalness, it has been conjectured that expanding the econometrician's information set might solve the non-fundamentalness problem.[1] The proposed test of this paper can be used to formally test if adding more information solves the non-fundamentalness problem. Specifically, I show that augmenting the baseline VAR model with a narrative variable that measure *news* about future government spending restores fundamentalness. Consequently, an econometrician can proceed with the identification strategy that she finds reasonable to recover the structural shocks.

The rest of the paper is organized as follows: Section 2 provides a formal statement of the fundamental representation and the testing problem. Section 3 introduces formally the test statistic based on the generalized spectrum. Section 4 examines the finite-sample performance of the test through some Monte Carlo simulation based on a DSGE model and an empirical application to the identification of government spending shocks. Section 5 concludes. The MATLAB code for implementing the test is available from the author upon request.

2 Characterization of non-fundamental VARMA representations

Consider the following d-variate zero mean VARMA(p,q) model in standard representation:

$$x_t = \sum_{i=1}^{p} \phi_i x_{t-i} + \xi_t + \sum_{j=1}^{q} \theta_j \xi_{t-j}$$

where ξ_t is a sequence of independent random vectors defined on some probability space $(\Omega, \mathcal{A}, \mathcal{P})$. The vectors x_t and ξ_t contain the d univariate time series: $x_t = [x_{1t}, x_{2t}, \cdots, x_{dt}]'$

[1]See for example, Giannone and Reichlin (2006) and Forni and Gambetti (2014).

and $\xi_t = [\xi_{1t}, \xi_{2t}, \cdots, \xi_{dt}]'$. We can also write the previous equation in lag operators:

$$\Phi(L)x_t = \Theta(L)\xi_t \ , \qquad t = 0, \pm 1, \pm 2, \cdots \tag{2.1}$$

where

$$\Phi(L) := I_d - \Phi_1 L - \cdots - \Phi_p L^p$$
$$\Theta(L) := I_d + \Theta_1 L + \cdots + \Theta_q L^q$$

are the AR and MA polynomials, respectively. Henceforth, $\mathbf{I}_d$ is the $d \times d$ identity matrix, $\Phi_p \neq 0$ and $\Theta_q \neq 0$ and L is the lag operator, i.e., $Lx_t = x_{t-1}$. The polynomials $\Phi(\cdot)$ and $\Theta(\cdot)$ have no common roots and neither of the roots is on the unit circle.

Although the solution of DSGE models can be reduced to a VARMA model by log-linearizing around their steady-state, VARMA models are rarely used to represent multivariate time series processes. Instead, due to their convenience, VAR models are widely employed as an approximation to the VARMA process and to recover the structural shocks. However, the VAR representation is admissible only under fundamentalness, also known as invertibility.[2]To begin with, lets define fundamentalness (see Rozanov (1967) and Alessi et al. (2011)).

Definition 2.1: An uncorrelated process $\{\xi_t\}$ is x_t-fundamental if $\mathcal{H}_t^\xi = \mathcal{H}_t^x$ for all $t \in \mathbb{Z}$, where $\mathcal{H}_t^\xi$ is the closed linear span of $\{\xi_s : s \leq t\}$. The process $\{\xi_t\}$ is non-fundamental if $\mathcal{H}_t^\xi \in \mathcal{H}_t^x$ and $\mathcal{H}_t^\xi \neq \mathcal{H}_t^x$, for at least one $t \in \mathbb{Z}$.

A VARMA process defined by (2.1) is said to be fundamental if and only if all the roots of $\det[\Theta(z)]$ lie outside the unit circle in the complex plane.[3] One can show that if non-fundamental representation is excluded by mistake, the true unobserved shocks will

[2]Fundamentalness is slightly different from invertibility, since invertibility requires that no roots of the MA component be on or inside the unit circle. In this framework, they are equivalent since unit root in the MA polynomial is ruled out.

[3]Similarly (2.1) is said to be causal if and only if all the roots of $\Phi(z)$ lie outside the unit circle in the complex plane. See Brockwell and Davis (1991), Theorems 3.1.1 and 3.1.2. Throughout, I assume that the model is causal.

be related to the Wold innovations through *Blaschke* matrices. *Blaschke* matrices are complex-valued filters which take the roots from inside to outside the unit circle, thus generates a fundamental representation from a non-fundamental one (Lippi and Reichlin, 1994). The following example illustrates the main ideas.

Example 2.1: Leeper et al. (2013) introduce foresight into a simple growth model. Assuming two-quarter fiscal foresight, the log-linearized equilibrium condition for capital is

$$(1 - \alpha L)k_t = -\kappa(L + \theta)\xi_{\tau,t} \tag{2.2}$$

where κ is a functions of the deep parameters of the model and $0 < \alpha < 1$ and $0 < \theta < 1$. However, fundamentalness is satisfied only if $|\theta| > 1$. The fact that more recent tax news are discounted heavier than older news makes model (2.2) non-fundamental. Imposing fundamentalness, the less informed econometrician incorrectly estimates the model

$$(1 - \alpha L)k_t = -\kappa(1 + \theta L)\epsilon_{\tau,t} \qquad |\theta| < 1$$

or in the autoregressive form

$$\frac{(1 - \alpha L)}{-\kappa(1 + \theta L)}k_t = \sum_{j=0}^{\infty} \gamma_j k_{t-j} = \epsilon_{\tau,t} \qquad |\theta| < 1$$

where γ_j is a function of deep parameters and $\epsilon_{\tau,t}$ is the Wold innovation[4], related to the true unobserved errors through *Blaschke* factor, $\epsilon_{\tau,t} = \left[\frac{L+\theta}{1+\theta L}\right]\xi_{\tau,t}.\square$

In practice, it is common to estimate a VAR instead of a VARMA, which makes detecting non-fundamentalness more complicated since the DGP has undergone a further approximation. To see this, suppose the true process is a non-fundamental ARMA process (2.1), but an econometrician incorrectly imposes fundamentalness assumption. One can show that the resulting process has a representation given by

$$\Phi(L)x_t = \tilde{\Theta}(L)\epsilon_t \tag{2.3}$$

[4] i.e., $\epsilon_t = k_t - \mathbb{L}[k_t|\mathcal{H}_t^k]$ where, $\mathbb{L}[k_t|\mathcal{H}_t^k]$ denotes the optimal linear predictor of k_t given its past.

where $\tilde{\Theta}(L)$ has the same order as $\Theta(L)$ but all its roots are outside the unit circle and $\{\epsilon_t\}$ are the Wold innovations related to the original innovations, $\{\xi_t\}$, through

$$\epsilon_t = \tilde{\Theta}^{-1}(L)\Theta(L)\xi_t \tag{2.4}$$

where $\tilde{\Theta}^{-1}(L)\Theta(L)$ is the *Blaschke* factor. Therefore, (2.3) can be written as a VAR(∞) form:

$$\tilde{\Theta}(L)^{-1}\Phi(L)x_t = \sum_{j=0}^{\infty} \gamma_j x_{t-j} = \epsilon_t \tag{2.5}$$

For estimation of such models it is necessary to approximate the infinite order lag structure by finite order VAR(p). In practice, the order p is often selected so that the residuals are approximately white noise. One can prove that if fundamentalness is imposed incorrectly, the Wold innovations (2.4) are still uncorrelated. Therefore, estimation methods based on second-order moment techniques do not identify non-fundamentalness. In order to deal with this identification problem the literature imposes fundamentalness by assumption.

In the non-Gaussian case, however, fundamental and non-fundamental models are distinguishable based on higher order cumulants (Lii and Rosenblatt, 1982). Using time-reversibility argument, Breidt and Davis (1992) proved that the Wold innovations from fitting an invertible ARMA model to a non-invertible one are *iid*, if and only if the error is non-Gaussian. Chen et al. (2012) extended this result to the multivariate case and proposed to test for serial dependence to detect non-fundamentalness. However, testing for the violation of the *iid* assumption of the Wold innovations is restrictive and may lead to rejection of the null of fundamentalness by mistake. The following is an example intended to highlight this point.

Example 2.2: Consider the ARCH process

$$\begin{aligned} &x_t = \xi_t \\ &\xi_t = h_t^{1/2} z_t \quad h_t = 0.43 + 0.57 z_{t-1}^2 \\ &z_t \sim iid\ N(0,1) \end{aligned}$$

Definition 2.1 trivially holds and therefore ξ_t is x_t-fundamental. However, ξ_t is an ARCH process and therefore a serial dependence test can incorrectly reject the null of fundamentalness.□

In this paper, I use the information available in the higher order moments of the true unobserved shocks to propose a new test which is robust to the failure of the *iid* assumption. Under some mild conditions stated in Assumption 1, I prove that if the model is non-fundamental, the Wold innovations are non-MD, i.e., non-linearly predictable despite being white noise.

Assumption 1. Let ξ_{jt} denote the jth element of the true unobserved shocks $\{\xi_t\}$. Then for all $j \in 1, \cdots, d$, ξ_{jt} is distributed with a non-Gaussian distribution such that $(a+1)th$ moment finite with $(a+1)th$ cumulant nonzero for some $a \geq 2$.

Proposition 2.1: Let Assumption 1 hold. The non-Gaussian VARMA model (2.1) is invertible if and only if the Wold innovations $\{\epsilon_t\}$ defined in (2.4), are unpredictable.

For the proof see Appendix A. Non-Gausianity assumption is needed to achieve identification. In fact, there is mounting evidence that emphasizes considering non-Gaussian distributions and other higher order time-varying moments (see e.g., Harvey and Siddique, 1999, 2000; Jondeau and Rockinger, 2003). Note that, no specific distributional assumption is needed.

3 Testing for non-fundamental representations

Under the null of fundamentalness $\xi_t(\theta_0) = \epsilon_t(\theta_0)$, which following Proposition 2.1 can be restated as

$$\mathbb{H}_0 : \epsilon_t(\theta_0) \text{ are MD (unpredictable) for some } \theta_0 \in \Xi \tag{3.1}$$

where $\theta_0 = vec\{\Phi_1, \cdots, \Phi_p, \Theta_1, \cdots, \Theta_q, \Sigma_\epsilon\}$, and $vec(.)$ denote an operator on a matrix which cascades the columns of the matrix from the left to the right and forms a column vector.

Testing (3.1) is not an easy task. Portmanteau tests proposed by Box and Pierce (1970) and Ljung and Box (1978) are not suitable to reflect the non-linear dependence structure. Moreover, $\{\epsilon_t\}$ is unobserved and residuals depend on a $\sqrt{T}$-consistent estimator for θ_0, which may cause the loss of the nuisance parameter-free property of the asymptotic distribution of the test statistics.

To overcome these problems and checking for unpredictability at all lags in the sample, I extend the generalized spectral test of Hong and Lee (2005) to the multivariate setting. Compared with the existing tests in the literature, this test has some advantages: first, with the frequency domain approach, one can allow infinite number of lags as the sample size increases; second, the test has a standard normal limiting distribution and parameter estimation uncertainty has no impact on the asymptotic distribution of the test statistics. The proposed test can also be used to test the martingale hypothesis in the multivariate setting for observed raw data without any modification.

My proposal for testing the MD property of the Wold innovations is based upon the generalized spectrum of Hong (1999):

$$f(\omega, u, v) \equiv \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \sigma_j(u, v) \exp(-ij\omega), \tag{3.2}$$

where $\omega \in [-\pi, \pi]$ is the frequency, $i \equiv \sqrt{-1}$, $(u, v) \in \mathbb{R}^d \times \mathbb{R}^d$, and

$$\sigma_j(u, v) = \operatorname{cov}\big(\exp(iu'\epsilon_t), \exp(iv'\epsilon_{t-|j|})\big), \quad j = 0, \pm 1, ...$$

where $\epsilon_t \equiv \epsilon_t(\theta)$. Note that $f(\omega, u, v)$ is a complex-valued scalar function, although ϵ_t is a $d \times 1$ vector. The function $f(\omega, u, v)$ captures any type of pairwise serial dependence in $\{\epsilon_t\}$, including that with zero autocorrelation function.

The generalized spectrum $f(\omega, u, v)$ is not suitable for testing (3.1), because it also captures the serial dependence in higher order moments. For example, $f(\omega, u, v)$ captures GARCH dependence, although the process could be a MD. However, just as the characteristic function can be differentiated to generate various moments of ϵ_t, $f(\omega, u, v)$ can be differentiated to capture the serial dependence in various moments. To capture (and only capture) the serial dependence in the conditional mean, one can use

$$f^{(0,1,0)}(\omega, u, v) \equiv \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \sigma_j^{(1,0)}(0, v) \exp(-ij\omega), \quad \omega \in [-\pi, \pi]$$

where

$$\sigma_j^{(1,0)}(0, v) \equiv \frac{\partial}{\partial u} \sigma_j(u, v)\big|_{u=0} = \operatorname{cov}\big(i\epsilon_t, \exp(iv'\epsilon_{t-|j|})\big)$$

is a $d \times 1$ vector. The measure $\sigma_j^{(1,0)}(0, v)$ checks whether the autoregression function $E(\epsilon_t|\epsilon_{t-j}) = 0$ at lag j is zero.[5]

In the present context, ϵ_t is not observed. Suppose we have T observations $\{x_t\}_{t=1}^T$ which are used to estimate the model and to obtain the estimated model residuals

$$\hat{\epsilon}_t \equiv \hat{\Theta}^{-1}(L)\hat{\Phi}(L)x_t \tag{3.3}$$

where $\hat{\theta}$ is a $\sqrt{T}$-consistent estimator for θ_0. Examples of $\hat{\theta}$ are conditional least squares and quasi-maximum likelihood estimator. We can estimate $f^{(0,1,0)}(\omega, 0, v)$ by a smoothed kernel estimator

$$\hat{f}^{(0,1,0)}(\omega, 0, v) \equiv \frac{1}{2\pi} \sum_{j=T-1}^{T-1} (1 - \frac{|j|}{T})^{1/2} k(j/h) \hat{\sigma}_j^{(1,0)}(0, v) \exp(-ij\omega), \quad \omega \in [-\pi, \pi] \tag{3.4}$$

where $\hat{\sigma}_j^{(1,0)}(0, v) = \frac{\partial}{\partial u} \hat{\sigma}_j(u, v)\big|_{u=0}$, $\hat{\sigma}_j(u, v) = \hat{\varphi}_j(u, v) - \hat{\varphi}_j(u, 0)\hat{\varphi}_j(0, v)$, and

$$\hat{\varphi}_j(u, v) = \frac{1}{T - |j|} \sum_{t=j+1}^{T} \exp(iu'\hat{\epsilon}_t + iv'\hat{\epsilon}_{t-|j|})$$

where $h \equiv h(T)$ is a bandwidth, and $k : \mathbb{R} \to [-1, 1]$ is a symmetric kernel. Examples of $k(\cdot)$ include the Bartlett, Daniell, Parzen and Quadratic spectral kernels. The factor $(1 - \frac{|j|}{T})^{1/2}$ is a finite-sample correction. The effect of this correction factor is to put less weight on very large lags, for which we have less sample information. It could be replaced by unity.

[5]The hypothesis of $E(\epsilon_t|I_{t-j}^{\epsilon}) = 0$ *a.s.* is not the same as the hypothesis of $E(\epsilon_t|\epsilon_{t-j}) = 0$ *a.s.* for all $j > 0$. The former checks all type of dependencies, whereas the latter one only captures pairwise dependencies. See Hong (1999) for more discussion on this.

Under $\mathbb{H}_0$, the generalized spectral derivative $f^{(0,1,0)}(\omega, 0, v)$ becomes a flat spectrum:

$$f_0^{(0,1,0)}(\omega, 0, v) \equiv \frac{1}{2\pi}\sigma_0^{(1,0)}(0, v), \quad \omega \in [-\pi, \pi]$$

which can be consistently estimated by

$$\hat{f}_0^{(0,1,0)}(\omega, 0, v) \equiv \frac{1}{2\pi}\hat{\sigma}_0^{(1,0)}(0, v), \quad \omega \in [-\pi, \pi].$$

The estimators $\hat{f}^{(0,1,0)}(\omega, 0, v)$ and $\hat{f}_0^{(0,1,0)}(\omega, 0, v)$ converge to the same limit under $\mathbb{H}_0$, and generally converge to different limits under $\mathbb{H}_1$. Thus, any significant divergence between them can be interpret as evidence of the violation of the MD property, and hence, of the non-fundamentalness of the process.

The test statistic, that is robust to conditional heteroscedasticity and other time-varying higher order conditional moments of unknown form, is given as follows:

$$\hat{M} \equiv \left[\sum_{j=1}^{T-1} k^2(j/h) T_j \int \left\|\hat{\sigma}_j^{(1,0)}(0, v)\right\|^2 d\mathcal{W}(v) - \hat{C}\right] \Big/ \sqrt{\hat{D}} \tag{3.5}$$

where $T_j = T - j$, $\mathcal{W}(v) = \prod_{c=1}^{d} W(v_c)$, $W : \mathbb{R} \to \mathbb{R}^+$ is a nondecreasing function that weight sets symmetric about zero equally, and the unspecified integrals are taken over the support of $\mathcal{W}(\cdot)$. Examples of $W(\cdot)$ include the CDF of any symmetric probability distribution, either discrete or continuous. $\hat{C}$ and $\hat{D}$ are estimate of the mean and the variance of $T \iint_{-\pi}^{\pi} \|\hat{f}^{(0,1,0)}(\omega, 0, v) - \hat{f}_0^{(0,1,0)}(\omega, 0, v)\|^2 d\omega d\mathcal{W}(v)$,

$$\hat{C} \equiv \sum_{j=1}^{T-1} k^2(j/p) \frac{1}{T-j} \sum_{t=j+1}^{T-1} \|\hat{\epsilon}_t\|^2 \int \left|\hat{\psi}_{t-j}(v)\right|^2 dW(v)$$

$$\hat{D} = 2\sum_{j=1}^{T-2}\sum_{l=1}^{T-2} k^2(j/p)k^2(l/p) \sum_{a=1}^{d}\sum_{b=a}^{d} \int\int \Bigg| \frac{1}{T - \max(j,l)} \times \sum_{t=\max(j,l)+1}^{T} \hat{\epsilon}_{at}\hat{\epsilon}'_{bt}\hat{\psi}_{t-j}(v)\hat{\psi}^*_{t-l}(v')\Bigg|^2 dW(v)dW(v')$$

where $\hat{\psi}_t(v) = \exp(iv'\hat{\epsilon}_t) - T^{-1}\sum_{t=1}^{T}\exp(iv'\hat{\epsilon}_t)$, and $\hat{s}^4 = \sum_{a,b=1}^{d}\left(T^{-1}\sum_{t=1}^{T}\hat{\epsilon}_{at}\hat{\epsilon}_{bt}\right)^2$.

To derive the limit distribution of the test, I need to impose some regularity conditions. Throughout, I use C to denote a generic bounded constant, $\|.\|$ the Euclidean norm, and A^* the complex conjugate of A.

Assumption A1. $\{x_t\}$ is a $d \times 1$ covariance stationary time series process, and ϵ_t are MD with $\mathrm{E}\|\epsilon_t^4\| \leq C$, where ϵ_t is Wold innovation from estimating an invertible model.

Assumption A2. For q sufficiently large, there exists a strictly stationary process $\{\epsilon_{q,t}\}$ measurable with respect to the sigma field generated by $\{\epsilon_{t-1}, \epsilon_{t-2}, \cdots, \epsilon_{t-q}\}$ s.t. as $q \to \infty$, $\epsilon_{q,t}$ is independent of $\{\epsilon_{t-q-1}, \epsilon_{t-q-2}, \cdots\}$ for each t, $\mathrm{E}[\epsilon_{q,t}|I_{t-1}] = 0$ a.s., $\mathrm{E}\|\epsilon_t - \epsilon_{q,t}\|^2 \leq Cq^{-\kappa}$ for some constant $\kappa \geq 1$, and $\mathrm{E}\|\epsilon_{q,t}\|^4 \leq C$ for all large q.

Assumption A3. The estimator $\hat{\theta}$ is such that $\sqrt{T}(\hat{\theta} - \theta^*) = O_P(1)$, where $\theta^* \equiv plim_{T\to\infty}\hat{\theta}$. Under $\mathbb{H}_0$, $\theta^* = \theta_0$.

Assumption A4. Let $\bar{x}_0 = (x_0; \cdots; x_{1-p}; \epsilon_0; \cdots; \epsilon_{1-q})$ be some assumed initial values. Then $\mathrm{E}\|\bar{x}_0^2\| < \infty$.

Assumption A5. $k : \mathbb{R} \to [-1, 1]$ is symmetric about 0, and is continuous at 0 and all points except a finite number of points, with $k(0) = 1$ and $|k(z)| \leq C|z|^{-b}$ as $z \to \infty$ for some $b > 1$.

Assumption A6. $W : \mathbb{R} \to \mathbb{R}^+$ is nondecreasing and weights sets symmetric about zero equally, with $\int \|v\|^4 dW(v) \leq C$.

Assumption A7. Define $\psi_t(v) \equiv \exp(iv\epsilon_t) - T^{-1}\sum_{t=1}^{T}\exp(iv\epsilon_t)$ and $\Sigma \equiv \mathrm{E}(\epsilon_t\epsilon_t')$. Then, $\{\frac{\partial \epsilon_t}{\partial \theta}, \epsilon_t\}$ is a strictly stationary process such that

(a) $\sum_{j=1}^{\infty}\|\mathrm{cov}[\frac{\partial \epsilon_t}{\partial \theta}, \psi_{t-j}(v)]\| \leq C$;

(b) $\sum_{j=1}^{\infty}\sup_{(u,v)\in\mathbb{R}^2}|\sigma_j(u,v)| \leq C$;

(c) $\sum_{j=1}^{\infty}\sum_{l=1}^{\infty}\sup_{(u,v)\in\mathbb{R}^2}\left\|\mathrm{E}[(\epsilon_t\epsilon_t' - \Sigma)\psi_{t-j}(u)\psi_{t-l}(v)]\right\| \leq C$;

(d) $\sum_{j=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}\sum_{\tau=-\infty}^{\infty}\sup_{v\in\mathbb{R}}\|\kappa_{j,l,\tau}(v)\| \leq C$, where $\kappa_{j,l,\tau}(v)$ is the fourth order cumulant of the joint distribution of the process $\{\frac{\partial\epsilon_t}{\partial\theta}, \psi_{t-j}(v), \frac{\partial\epsilon_{t-l}}{\partial\theta}, \psi^*_{t-\tau}(v)\}$.

Assumption A8. $\sum_{j=1}^{\infty}\sup_{v\in\mathbb{R}}\|\sigma_j^{(1,0)}(0,v)\| \leq C$.

Assumption A1 is a regularity condition on the data generating process (DGP) $\{x_t\}$. Assumption A2 is required only under $\mathbb{H}_0$, which states that the MD $\{\epsilon_t\}$ can be approximated by a q-dependent MD process $\{\epsilon_t\}$ arbitrarily well when q is sufficiently large. Because $\{\epsilon_t\}$ is a MD, Assumption A2 essentially imposes restrictions on the serial dependence in higher order moments of $\{\epsilon_t\}$. It covers GARCH and stochastic volatility processes as special cases; see *e.g.* Hong and Lee (2005). Assumption A3 requires a $\sqrt{T}$-consistent estimator $\hat{\theta}$, such as conditional least squares estimator or a conditional quasi-maximum likelihood estimator.

Assumption A4 is a start-up value condition. It ensures that the impact of initial values assumed in the observed information set is asymptotically negligible. Assumption A5 is a regularity condition on the kernel $k(.)$. It includes all commonly used kernels in practice. For kernels with bounded support, such as the Bartlett and Parzen kernels, we have $b = \infty$: For kernels with unbounded support, b is some finite positive real number. Assumption A6 is a condition on the weighting function $W(.)$ for the transform parameter v. It is satisfied by the CDF of any symmetric continuous distribution with a finite fourth moment. Assumption A7 provides some covariance and fourth order cumulant conditions on $\{\frac{\partial\epsilon_{t-1}}{\partial\theta}, \epsilon_t\}$, which restricts the degree of serial dependence in $\{\frac{\partial\epsilon_{t-1}}{\partial\theta}, \epsilon_t\}$. Finally, Assumption A8 impose a condition on the serial dependence in $\{\epsilon_t\}$. The asymptotic properties of the test statistic are stated in the following theorem. The proof is similar to the univariate case of Hong and Lee (2005), and for the sake of space is only provided in the online appendix.

Proposition 4.1: Let $h = cT^{\lambda}$ for $0 < \lambda < (3 + \frac{1}{4b-2})^{-1}$, where b is the same as assumption A5, and $0 < c < \infty$. Then:

(a) Under Assumptions A1-A7 and $\mathbb{H}_0$, $\hat{M} \xrightarrow{d} N(0,1)$.

(b) Under Assumptions A1-A8 and $\mathbb{H}_1$, $\lim_{T\to\infty} P[\hat{M} > C(T)] = 1$ for any sequence $C(T) = o(T/h^{1/2})$.

Under the null, $\hat{M}$ has a simple standard normal distribution. Under the alternative hypothesis, $E(\epsilon_t|\epsilon_{t-j}) \neq 0$ a.s., at some lag $j > 0$. Then we have $\int \|\sigma_j^{(1,0)}(0,v)\|^2 d\mathcal{W}(v) > 0$ for any weighting function $\mathcal{W}(\cdot)$ that is positive, monotonically increasing and continuous, with unbounded support on $\mathbb{R}$. Therefore, $\hat{M}$ has asymptotic unit power at any given significance level.

An important feature of $\hat{M}$ is that the use of the estimated residuals $\{\hat{\epsilon}_t\}$ in place of the true errors $\{\epsilon_t\}$ has no impact on the limit distribution of $\hat{M}$. The reason is that the convergence rate of the parametric parameter estimator $\hat{\theta}$ to θ_0 is faster than that of the nonparametric kernel estimator $\hat{f}^{(0,1,0)}(w,0,v)$ to $f^{(0,1,0)}(w,0,v)$. Consequently, the limit distribution of $\hat{M}$ is solely determined by $\hat{f}^{(0,1,0)}(w,0,v)$, and replacing θ_0 by $\hat{\theta}$ has no impact asymptotically.

4 Monte Carlo evidence and empirical application

4.1 Simulation study

In this section I examine the finite sample performance of the proposed test based on artificial data generated from the DSGE model with fiscal foresight of Leeper et al. (2013). The model is characterized by a representative household that maximizes expected log utility,

$$E_0 \sum_{t=0}^{\infty} \beta^t \log(C_t)$$

$$s.t. \qquad C_t + K_t + T_t \leq (1-\tau_t) A_t K_{t-1}^{\alpha}$$

where $C_t,\ K_t,\ Y_t,\ T_t$, and τ_t denote time$-t$ consumption, capital, output, lump-sum taxes, and the income tax rate, respectively, and A_t is an exogenous technology shock. The parameters satisfy $0 < \alpha < 1$, $0 < \beta < 1$. The government sets the tax rate according to $T_t = \tau_t Y_t$, and labor is supplied inelastically. Let A and τ_k denote the steady states values of technology and the tax rate. The log-linearized equilibrium condition for the capital and the tax rate is given by the following bivariate VARMA model

$$\hat{\tau}_t = \Psi(L)\xi_{\tau,t}$$

$$k_t = \alpha k_{t-1} + \xi_{a,t} - \frac{\tau(1-\theta)}{1-\tau}\sum_{k=0}^{\infty}\theta^k E_t\hat{\tau}_{t+k+1}$$

where $\theta = \alpha\beta\frac{1-\tau_y}{1-\tau_k}$ and the lower case letters denote percentage deviations from steady state values, $k_t = \log(K_t) - \log(K)$, $a_t = \log(A_t) - \log(A)$, and $\hat{\tau}_t = \log(\tau_t) - \log(\tau)$.

To model foresight, I assume the tax rate evolves as

$$\hat{\tau}_t = \sum_{j=0}^{J}\psi_j\xi_{\tau,t-j} = \Psi(L)\xi_{\tau,t} \tag{4.1}$$

where $\sum_{j=0}^{J}\psi_j = 1$, and $\psi_j \in [0,1]$ determines the relative weight of the shock at time j. Table 1 presents different processes for the tax rate that are used for the Monte Carlo study. DGP1-DGP4 are examples of fundamental processes, and therefore VAR should give a reasonable approximation. DGP5-DGP8 are examples of non-fundamental processes and therefore, they do not admit VAR representation mapping economic shocks to a vector of observable variables and its lags. DGP8 is an example of non-fundamental processes with roots zero, which are commonly used in the literature with news shocks (see e.g., Schmitt-Grohé and Uribe (2012) and Forni et al. (2014)). Although Proposition 2.1 rules out processes that have a roots equal to zero, it would be interesting to see how the test performs. It is also interesting to study the performance of the tests when the roots are very close to the unit root. I choose the weights of DGP2 and DGP6 so that the roots are very close to the unit root.

For the simulation exercise, I generate series for the capital and the tax rate setting $\alpha = 0.36$, $\beta = 0.99$, and $\tau = 0.25$, as in Leeper et al. (2013). The structural shocks $\xi_{a,t}$ and $\xi_{\tau,t}$ are generated as centered *iid lognormal*$(0,1)$, mutually independent at all leads and lags. To examine why it is important to take into account the impact of conditional moments in testing $\mathbb{H}_0$, I also consider a GARCH process for $\xi_{a,t} = \sigma_t^{\frac{1}{2}}z_t, \quad \sigma_t^2 =$

$0.001 + 0.09\xi_{t-1}^2 + 0.9\sigma_{t-1}^2$ and $\xi_{\tau,t} \sim iid\ lognorm(0,1)$. A similar GARCH process is used by Escanciano and Velasco (2006).[6]

For the sake of comparison, I also report the result of the test proposed by Chen et al. (2012) who consider the stronger null hypothesis that the errors are serially independent. Their proposed test statistic to check for serial dependence of the residuals is of the form

$$\hat{Q} \equiv \left[\sum_{j=1}^{T-1} k^2(j/h) T_j \iint \left|\hat{\sigma}_j(u,v)\right|^2 d\mathcal{W}(u) d\mathcal{W}(v) - \hat{C}_q \right] \Big/ \sqrt{\hat{D}_q}$$

where

$$\hat{C}_q = \sum_{j=1}^{T-1} k^2(j/h) \left[\int \hat{\sigma}_0(v,-v) d\mathcal{W}(v) \right]^2$$

$$\hat{D}_q = 2 \sum_{j=1}^{T-2} k^4(j/h) \left[\int |\hat{\sigma}_0(u,v)|^2 d\mathcal{W}(u) d\mathcal{W}(v) \right]^2$$

which also has an asymptotic standard normal null distribution.

Some comments are in order. First, both $\hat{M}$ and $\hat{Q}$ involve $d-$ and $2d-$ dimensional numerical integration, which can be computationally cumbersome when d is large. In practice, one may approximate the integrals by choosing a finite number of grid points symmetric about zero or generate a finite number of points drawn from the uniform distribution on $[-1,1]^d$. Alternatively, for some weighting functions there is a closed form expression for the test statistics. In this paper, I use a closed form solution obtained by choosing $d\mathcal{W}(\cdot)$ as the $d-$dimensional Gaussian CDF.

Second, a practical issue in implementing the tests is the choice of the bandwidth parameter $\hat{h}$. Following Hong and Lee (2005), one can choose a data-driven bandwidth $\hat{h} = \hat{c}_0 T^{\frac{1}{2q+1}}$ via the plug-in method, which lets data themselves determine an appropriate lag.[7] The data-driven bandwidth $\hat{c}_0$, involves the choice of a preliminary bandwidth $\bar{h}$, which can be fixed or grow with the sample size T. Applying the data-driven method to

[6] As a robustness check, I examined many combinations of alternative volatility forms and found results that are consistent with those of Table 2.

[7] q is called the characteristic exponent of $k(.)$. For Parzen, Daniell and quadratic spectral (QS) kernels, $q = 2$.

Table 1: Information Flow Processes

Process	Description	Coefficients	Roots
		Size	
DGP1	No foresight	$\psi_0 = 1$	–
DGP2	1-qtr concentrated news	$\psi_0 = 0.51, \psi_1 = 0.49$	$z = 1.04$
DGP3	1-qtr concentrated news	$\psi_0 = 0.8, \psi_1 = 0.2$	$z = 4$
DGP4	2-qtr concentrated news	$\psi_0 = 0.8, \psi_1 = 0.1, \psi_2 = 0.1$	$z_{1,2} = 2.83$
		Power	
DGP5	1-qtr concentrated news	$\psi_0 = 0.49, \psi_1 = 0.51$	$z = 0.96$
DGP6	1-qtr concentrated news	$\psi_0 = 0.2, \psi_1 = 0.8$	$z = 0.25$
DGP7	2-qtr concentrated news	$\psi_0 = 0.1, \psi_1 = 0.1, \psi_2 = 0.8$	$z_{1,2} = 0.35$
DGP8	2-qtr perfect foresight	$\psi_2 = 1$	$z_{1,2} = 0$

Note: (1) Coefficient settings in tax rule (4.1); (2) z's are the roots of the determinant of the moving average polynomial.

choose the bandwidth, while considering a wide range of the bandwidth, $\bar{h} \in \{2, \cdots, 10\}$, the simulation results show that the tests are not sensitive to the choice of preliminary bandwidth. For the sake of space, I only report the results for $\bar{h} = 3, 6$ and 9, using the Parzen kernel. Simulations suggest that the choice of $k(\cdot)$ has little impact on both the level and the power of the tests.

I estimate a VAR(p) based on a sample size of 250 which is about the size of most postwar data sets. The number of Monte Carlo replication is 500. I also throw away the first 1000 observations for removing initial conditions effects on the simulations. Although a non-fundamental model does not admit a VAR representation, it is interesting to know if considering larger number of the lag order of the estimated VAR will solve the non-fundamentalness problem. Therefore, I consider lag orders 4 and 8.

Table 2 reports the rejection rates of the tests at the 10% and 5% levels. The simulation results show that $\hat{M}$ under-rejects $\mathbb{H}_0$. Similar under-rejection has been reported by Hong and Lee (2005) for the univariate version of $\hat{M}$. These authors argue that the under-rejection is due to the parameter estimation uncertainty in the finite-sample. Estimating too many parameters under the null could be one reason that $\hat{M}$ under-rejects the null hypothesis. Intuitively, the asymptotic standard normal distribution only approximates the small sample distribution of the test statistic under the null hypothesis, and $T = 250$ is rather small. The fact that the test under-rejects the null hypothesis is not harmful, if power keeps high.

As can be seen from Table 2, $\hat{Q}$ does not control the size, even under the *iid* assumption. The performance is worst when one of the roots is close to the unit circle. The rejection of the null hypothesis of serial independence can be due to the truncation error. Theoretically, the truncation error associated with the estimation of a finite order VAR(p) which only approximates the exact infinite order VAR representation is expected to be small. However, it might be the case that the lag order p necessary to recover the structural shock maybe very large, and therefore the errors after truncation might be dependent even under the invertibility assumption; see also Ravenna (2007) and Soccorsi (2016). As can be seen form Panel B of Table 2, considering larger number of lags does not solve this problem.

Table 3 shows that $\hat{M}$ has good power against the alternative hypothesis, except for the case that the root is very close to the unit root and deteriorates rapidly when I increase the lag order to $p = 8$. The power of the test decrease slightly when I increase the lag order. However, considering larger lags does not solve the non-fundamentalness problem. The finding that choosing different lag order does not solve the problem is in accordance with the fact that if a model is non-invertible, we can not recover the true shocks even if we include infinite lags. In general, $\hat{Q}$ has better power properties, at the cost of not controlling the size.

4.2 Empirical application

As an empirical application, I focus on the dynamic effects of government spending shocks on economic activity in the United States. When the Great Recession severely hit the economies of most of the countries, it has been argued that fiscal policy should be the primary tool for the economy to recover. Yet there is a sharp conflict over the efficacy of discretionary fiscal policy. In particular, the effects of counter-cyclical policies such as increased government spending is controversial due to the fact that it is extremely difficult to isolate the exogenous effect of these policies on GDP.

Table 2: Testing for Fundamentalness: Size

		IID								GARCH							
		DGP1		DGP2		DGP3		DGP4		DGP1		DGP2		DGP3		DGP4	
		10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
Panel A: $p = 4$																	
$\bar{h} = 3$	$\hat{M}$	5.8	2.4	5.8	2.4	5.2	2.0	4.8	2.2	4.8	2.6	6.7	3.4	5.2	2.2	5.2	3.0
	$\hat{Q}$	27.2	21.2	66.6	52.0	26.2	19.0	25.0	19.8	46.2	38.6	77.4	68.0	40.2	33.0	41.0	34.6
$\bar{h} = 6$	$\hat{M}$	6.0	3.2	6.6	3.8	6.6	3.4	6.8	3.6	6.6	3.4	7.8	4.0	6.2	3.0	6.2	3.0
	$\hat{Q}$	28.0	22.4	70.8	58.4	27.6	19.8	27.0	20.8	48.8	41.8	79.8	72.8	43.4	35.0	44.0	36.2
$\bar{h} = 9$	$\hat{M}$	6.8	3.4	6.0	3.2	6.6	2.4	6.8	2.6	6.4	2.2	6.2	3.2	6.2	2.2	6.2	2.0
	$\hat{Q}$	30.4	23.2	72.2	59.8	28.4	20.8	28.6	21.4	50.2	43.4	82.6	75.4	45.2	36.8	47.0	38.8
Panel B: $p = 8$																	
$\bar{h} = 3$	$\hat{M}$	5.2	2.0	5.4	2.6	5.6	2.4	5.2	2.2	6.0	2.4	5.8	2.6	5.6	2.6	6.2	3.2
	$\hat{Q}$	24.8	17.4	42.4	30.8	24.4	17.2	24.8	17.4	45.6	38.8	40.0	32.4	40.4	33.4	42.4	36.4
$\bar{h} = 6$	$\hat{M}$	6.2	3.0	5.8	2.2	5.4	2.2	5.2	2.0	6.4	2.6	6.0	3.0	6.6	3.2	6.2	3.2
	$\hat{Q}$	25.2	19.4	44.6	34.2	25.0	18.4	26.8	18.4	40.6	40.2	42.2	35.0	42.8	34.8	44.4	37.4
$\bar{h} = 9$	$\hat{M}$	6.0	3.2	5.6	3.2	6.4	3.2	6.0	3.0	6.0	3.4	6.6	2.8	6.2	3.2	5.2	2.2
	$\hat{Q}$	26.0	20.2	48.2	37.2	26.4	19.0	27.0	19.2	48.4	41.2	44.2	36.4	45.6	36.6	47.2	38.8

Notes: (1) $\hat{M}$ is the multivariate martingale test; (2) $\hat{Q}$ is the multivariate independence test proposed by Chen et al. (2012); (3) $\bar{h}$ is the preliminary lag order used in a plug-in method to select a data-driven lag order; (4) The number of Monte Carlo replication is 500; (5) Sample size is 250.

Table 3: Testing for Fundamentalness: Power

		IID								GARCH							
		DGP5		DGP6		DGP7		DGP8		DGP5		DGP6		DGP7		DGP8	
		10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
Panel A: $p = 4$																	
$\bar{h} = 3$	$\hat{M}$	30.4	21.6	79.2	73.6	88.6	83.3	94.8	93.0	21.4	14.2	99.2	98.4	91.4	83.0	100	99.8
	$\hat{Q}$	93.4	89.6	99.6	99.6	94.6	92.8	96.8	94.0	89.2	82.0	100	100	100	100	100	100
$\bar{h} = 6$	$\hat{M}$	28.6	19.4	74.2	65.8	85.0	78.6	92.6	90.2	21.2	13.2	98.6	97.6	86.8	79.8	100	99.8
	$\hat{Q}$	94.2	90.2	99.6	99.4	95.2	93.2	96.8	94.4	91.0	84.4	100	100	100	100	100	100
$\bar{h} = 9$	$\hat{M}$	26.4	18.4	66.8	57.8	79.4	71.8	90.4	86.2	21.2	13.0	97.4	96.2	83.6	76.2	100	99.8
	$\hat{Q}$	94.8	91.4	99.6	99.2	96.0	94.4	96.8	95.4	92.2	87.8	100	100	100	100	100	100
Panel B: $p = 8$																	
$\bar{h} = 3$	$\hat{M}$	16.2	10.0	73.6	65.6	84.2	78.8	92.6	89.0	6.2	4.8	98.8	96.0	90.6	81.4	99.2	98.8
	$\hat{Q}$	75.4	66.6	99.6	99.0	91.6	88.6	97.0	94.6	60.4	52.6	100	99.8	100	100	100	100
$\bar{h} = 6$	$\hat{M}$	15.2	9.8	67.6	57.8	79.0	70.6	88.2	81.8	5.4	3.8	99.2	98.4	86.4	77.2	100	99.8
	$\hat{Q}$	77.4	69.2	99.6	99.2	92.8	89.6	97.2	94.8	63.2	54.4	99.8	99.8	100	100	100	100
$\bar{h} = 9$	$\hat{M}$	13.4	9.0	59.6	51.4	73.6	63.8	83.0	82.2	5.2	3.4	93.8	92.2	84.8	75.6	98.8	97.6
	$\hat{Q}$	79.8	72.2	99.6	99.2	92.8	90.4	97.2	95.2	65.4	57.2	99.8	99.8	100	100	100	100

Notes: (1) $\hat{M}$ is the multivariate martingale test; (2) $\hat{Q}$ is the multivariate independence test proposed by Chen et al. (2012); (3) $\bar{h}$ is the preliminary lag order used in a plug-in method to select a data-driven lag order; (4) The number of Monte Carlo replication is 500; (5) Sample size is 250.

Using VAR techniques, Blanchard and Perotti (2002) find moderate estimates of government spending output multipliers, an increase in consumption and the real wages (see also, Galí et al., 2007; Mountford and Uhlig, 2009). In contrast, Ramey (2011) argue that big increases in military spending are anticipated several quarters before they actually occur. Leeper et al. (2013) argue that fiscal foresight can create non-fundamentalness and therefore econometric methods using VAR models can not recover the correct structural shocks and impulse response functions.

To check whether fiscal foresight plays an important role in measuring the government spending shocks, I apply the test to the VAR specification standard in the empirical fiscal policy literature. To this end, suppose an economy is represented by a VMA model

$$x_t = \Gamma(L)\xi_t \tag{4.2}$$

where x_t consists of variables of interest and $\Gamma(L)$ is a polynomial in the lag operator. For the baseline specification, I include log real per capita quantities of total government spending(G), GDP (Y), Barro and Redlick (2011) tax rate (T), and the three-month T-bill rate($TB3$). This set of variables is similar to the ones used recently by Ramey (2011), covering the period 1948:I-2008:IV, and is available on Valerie Ramey's website.

Obtaining structural shocks from a VAR involves two steps: first, impose invertibility on (4.2) and construct a reduced form VAR model

$$\Pi(L)x_t = \xi_t \tag{4.3}$$

where $\Pi(L)$ is an autoregressive polynomial in the lag operator. Wold innovations can be recovered from estimating (4.3). Second, structural disturbances are identified from the reduced-form errors, imposing some identifying restrictions derived from economic theory or using a standard Choleski decomposition.

To apply the test, I only need model residuals from the first step and no identification strategy is required. This is consistent with the theory: no identification scheme is valid if the VAR is non-fundamental. Following Ramey (2011), I specify the VAR in levels, with a quadratic time trend and four lags included. Panel A of Table 4 reports the p-values of the tests applied to the residuals of this model.

Applying the tests to the residuals obtained from VAR, one observes that both $\hat{M}$ and $\hat{Q}$ reject the null of fundamentalness at the 10% level for the baseline specification. This implies that based on the results of the tests, given the data and variables selected in the baseline model, the impulse responses from SVAR approach appears not to be reliable.

Ramey (2011) argues that many shocks identified from a SVAR are anticipated changes in defense spending, which accounts for almost all of the volatility of government spending. Motivated by the importance of measuring anticipation, Ramey uses narrative evidence to construct a new variable, which measures the expected discounted value of government spending changes. Augmenting the baseline model with this narrative variable, Ramey finds very different effects of government spending on economic activities, and conjectures that this new narrative variable might solve the non-fundamentalness problem.

Table 4: Testing for Fundamentalness

	Parzen			Daniell			QS		
	$\bar{h}=3$	$\bar{h}=6$	$\bar{h}=9$	$\bar{h}=3$	$\bar{h}=6$	$\bar{h}=9$	$\bar{h}=3$	$\bar{h}=6$	$\bar{h}=9$
Panel A: Baseline = $[G,\ Y,\ T,\ TB3]$									
$\hat{M}$	0.049	0.062	0.091	0.031	0.061	0.085	0.039	0.069	0.091
$\hat{Q}$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Panel B: $News$−augmented = $[News,\ G,\ Y,\ T,\ TB3]$									
$\hat{M}$	0.387	0.389	0.377	0.352	0.371	0.395	0.379	0.392	0.394
$\hat{Q}$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Notes: (1) p-values for the null hypothesis that the structural model is fundamental; (2) $\hat{M}$ is the multivariate martingale test; (3) $\hat{Q}$ is the multivariate independence test proposed by Chen et al. (2012); (4) $\bar{h}$ is the preliminary lag order used in a plug-in method to select a data-driven lag order.

My proposal can be used to formally test if adding more information solves the non-fundamentalness problem. Panel B of Table 4 reports the p-values for the null of fundamentalness for the $\hat{M}$ and $\hat{Q}$, which suggest that we fails to reject the null for the *news*-augmented model. This implies that based on the results of the tests, the SVAR model augmented with the *news* variable is fundamental, and the impulse responses appear to be reliable. In contrast, serial dependence test, $\hat{Q}$, rejects the null of fundamentalness at 5% level for the *news*-augmented model. As discussed in the simulation study, this could be due to the fact that the $\hat{Q}$ test over-rejects the null hypothesis.

5 Conclusions

This paper provides a new theoretical and empirical tool for testing fundamentalness assumption of macroeconomic models. I convert the fundamentalness testing problem into one of testing the unpredictability of the Wold innovations. To test for the unpredictability of the innovations, I extend the generalized spectral density test of Hong and Lee (2005) to the multivariate case. The proposed test is simple to apply since it only needs model residual as input and has a convenient asymptotic standard normal null distribution. In addition, the test is robust to the failure of the independence assumption and does not need information outside of the specified model to check for fundamentalness. The Monte Carlo study based on a DSGE model with fiscal foresight exhibits a satisfactory finite-sample performance of the proposed test. I also show that choosing different lag orders when estimating a VAR model does not solve the invertibility problem, which is in accordance with the fact that if a model is non-invertible, we can not recover the true shocks even if we include infinite lags.

Furthermore, an empirical application to the identification of government spending shocks illustrates how to use the proposed test to a variety of empirical problems. I find that standard SVAR models commonly employed in the macroeconomics literature are non-fundamental. If the null hypothesis is rejected, it has been conjectured that expanding the econometrician's information set may restore the fundamentalness. The proposed test can be used to formally check if adding more information solves the non-fundamentalness problem. In the empirical application, I show that augmenting a standard VAR model with a narrative variable that measure anticipations solves the non-fundamentalness problem.

Appendix

I first prove Lemma 1, which is an extension of Theorem 5.4.1 Rosenblatt (2000), by dropping the identically distribution assumption. In Lemma 2, I use Lemma 1 to prove the univariate case of Proposition 2.1, and then show that under Assumption 1 the multivariate case can be reduced to the univariate case.

Lemma 1: Consider a univariate MA process obtained from (2.4)

$$\epsilon_t = \sum_{k=0}^{\infty} \gamma_k \xi_{t-k}, \quad \gamma_k = 0 \quad \forall k < 0 \tag{A.1}$$

and let $\phi^t(\tau)$ denote the characteristic function of ξ_t and $\phi^t_{\tau_0}(\cdot) = \frac{\partial \phi^t(\cdot)}{\partial \tau_0}$. Then linearity of the best predictor in mean square implies that

$$\sum_{k=-\infty}^{\infty} \big(\gamma_k - \sum_{l=1}^{\infty} \mathbf{b}_l \gamma_{k-l}\big) h^{t-k}\big(\sum_{l=1}^{\infty} \tau_l \gamma_{k-l}\big) = 0 \tag{A.2}$$

where $h^t(\vartheta) = \frac{\phi^t_{\tau_0}(\vartheta)}{\phi^t(\vartheta)}$ and $\mathbf{b}_l$'s are the coefficients of the best linear predictor of ϵ_t in mean square

$$\epsilon^*_t = \sum_{l=1}^{\infty} \mathbf{b}_l \epsilon_{t-l} \tag{A.3}$$

Proof of Lemma 1: The joint characteristic function of $\{\epsilon_{t-j}, j \geq 0\}$ is given by

$$\begin{aligned} \eta^t(\tau_0, \tau_1, \cdots, \tau_p, \cdots) &= E\Big\{ \exp\big(i \sum_{l=0}^{\infty} \tau_l \xi_{t-l}\big)\Big\} \\ &= \prod_{k=-\infty}^{\infty} \phi^{t-k}\big(\sum_{l=0}^{\infty} \tau_l \gamma_{t-l}\big) \end{aligned} \tag{A.4}$$

while the joint characteristic function of $\{\epsilon_{t-j}, j \geq 1\}$ is

$$\tilde{\eta}^t(\tau_1, \cdots, \tau_p, \cdots) = \prod_{k=-\infty}^{\infty} \phi^{t-k}\big(\sum_{l=1}^{\infty} \tau_l \gamma_{t-l}\big) \tag{A.5}$$

Differentiating $\eta^t(\tau_0, \tau_1, \cdots, \tau_p, \cdots)$ w.r.t. τ_0 we have

$$\begin{aligned}
&\frac{\partial}{\partial \tau_0}\eta^t(\tau_0, \tau_1, \cdots, \tau_p, \cdots)|_{\tau_0=0} = \eta^t_{\tau_0}(0, \tau_1, \cdots, \tau_p, \cdots) \\
&\quad = \int i\epsilon_t \ \exp(i\sum_{l=1}^{\infty}\tau_l\epsilon_{t-l}) \ dF^t(\epsilon_t, \epsilon_{t-1}, \cdots, \epsilon_{t-p}, \cdots) \\
&\quad = i\int E[\epsilon_t|\epsilon_{t-s}, s>0] \ \exp(i\sum_{l=1}^{\infty}\tau_l\epsilon_{t-l}) \ dF^t(\epsilon_{t-1}, \cdots, \epsilon_{t-p}, \cdots)
\end{aligned} \tag{A.6}$$

where $F^t(\epsilon_t, \epsilon_{t-1}, \cdots, \epsilon_{t-p}, \cdots)$ is the joint cumulative distribution function of $\epsilon_{t-j}, j \geq 0$. Also by differentiating the logarithm of (A.4) w.r.t. τ_0 we get:

$$\frac{\eta^t_{\tau_0}(0, \tau_1, \cdots, \tau_p, \cdots)}{\eta^t(0, \tau_1, \cdots, \tau_p, \cdots)} = \sum_{k=-\infty}^{\infty}\gamma_k h^{t-k}(\sum_{l=1}^{\infty}\tau_l\gamma_{k-l}). \tag{A.7}$$

Similarly, differentiating the logarithm of $\tilde{\eta}^t(\tau_1, \cdots, \tau_p, \cdots)$ w.r.t. τ_j, $j = 1, 2, \cdots$, we have

$$\frac{\partial}{\partial \tau_j} log \ \tilde{\eta}^t(\tau_1, \cdots, \tau_p, \cdots) = \sum_{k=-\infty}^{\infty}\gamma_{k-j}h^{t-k}(\sum_{l=1}^{\infty}\tau_l\gamma_{k-l}), \quad j = 1, 2, \cdots \tag{A.8}$$

If the best predictor in mean square is linear we must have

$$\eta^t_{\tau_0}(0, \tau_1, \cdots) = \sum_{k=1}^{\infty}\mathbf{b}_k\tilde{\eta}^t_{\tau_k}(\tau_1, \tau_2, \cdots) \tag{A.9}$$

which implies

$$\sum_{k=-\infty}^{\infty}\big(\gamma_k - \sum_{l=1}^{\infty}\mathbf{b}_l\gamma_{k-l}\big)h^{t-k}(\sum_{l=1}^{\infty}\tau_l\gamma_{k-l}) = 0. \tag{A.10}$$

□

Lemma 2: Let Assumption 1 hold. The univariate non-Gaussian ARMA model (2.1) is invertible if and only if the Wold innovations $\{\epsilon_t\}$ are MDS.

Proof of Lemma 2: A standard result for ARMA processes is that any ARMA(p, q) process $\{x_t\}$ which is non-invertible with respect to the noise sequence $\{\xi_t\}$ can also be modeled as an invertible ARMA(p, q) with respect to a new noise sequence $\{\epsilon_t\}$ defined by[8]

$$\epsilon_t = \frac{\prod\limits_{r_\Theta < i \leq q} (1 - b_i^{-1} L)}{\prod\limits_{r_B < i \leq q} (1 - b_i L)} \xi_t, \tag{A.11}$$

where $b_i, i = 0, \cdots, q - r_\Theta$ are the roots of the MA polynomial inside the unit circle. Above can be written as:

$$\sum_{i=0}^{q-r_\Theta} \alpha_i \epsilon_{t-i} = \sum_{i=0}^{q-r_\Theta} \beta_i \xi_{t-i}, \tag{A.12}$$

where $q - r_\Theta$ is the number of roots inside the unit circle.

Let $y_t = \sum_{i=0}^{q-r_\Theta} \alpha_i \epsilon_{t-i}$. Then (A.12) can be written as:

$$y_t = \sum_{i=0}^{q-r_\Theta} \beta_i \xi_{t-i}. \tag{A.13}$$

Because y_t in a non-invertible MA of order $(q - r_\Theta)$, Lemma 1 and Corollary 5.4.3 of Rosenblatt (2000) implies that the best one-step predictor of y_t is non-linear, i.e., $\mathrm{E}[y_t | y_{t-s}, s \geq 1]$ is non-linear. On the other hand, y_t is causal since all the roots of $\prod\limits_{r_\Theta < i \leq q} (1 - b_i L)$ are outside the unit circle. Therefore, the $\sigma-$algebras $\sigma(\epsilon_{t-s}, s \geq 1)$ and $\sigma(y_{t-s}, s \geq 1)$ coincide, and

$$\begin{aligned}
\mathrm{E}[y_t | y_{t-s}, s \geq 1] &= \mathrm{E}[y_t | \epsilon_{t-s}, s \geq 1] \quad a.s. \\
&= \mathrm{E}[\epsilon_t - \alpha_1 \epsilon_{t-1} - \cdots - \alpha_{q-r_\Theta} \epsilon_{t-q-r_\Theta} | \epsilon_{t-s}, s \geq 1] \quad a.s. \\
&= \mathrm{E}[\epsilon_t | \epsilon_{t-s}, s \geq 1] - \alpha_1 \epsilon_{t-1} - \cdots - \alpha_{q-r_\Theta} \epsilon_{t-q-r_\Theta} \quad a.s.
\end{aligned} \tag{A.14}$$

If ϵ_t were a MD, i.e. $\mathrm{E}[\epsilon_t | \epsilon_{t-s}, s \geq 1] = 0$, then

$$\mathrm{E}[y_t | y_{t-s}, s \geq 1] = -\alpha_1 \epsilon_{t-1} - \cdots - \alpha_{q-r_\Theta} \epsilon_{t-q-r_\Theta} \quad a.s. \tag{A.15}$$

which is linear -a clear contradiction- and therefore ϵ_t can not be a MD.□

Proof of Proposition 2.1: It is clear that if (2.1) is invertible, then $\{\epsilon_t\} \equiv \{\xi_t\}$ is unpredictable. We want to prove the reciprocal, that is if (2.1) is noninvertible then $\{\epsilon_t\}$

[8]See Brockwell and Davis (1991), page 103.

is not MDS. The proof in the univariate case follows from Lemma 2. To prove the multivariate case, let

$$\tilde{\Theta}^{-1}(L)\Theta(L) = A(L) = \Big[A_1(L)_{d\times 1} \;\; A_2(L)_{d\times(d-1)}\Big],$$

and define

$$\big[\epsilon_t \;\; M\big]_{d\times d} := \Big[A(L)\xi_t \;\; A_2(1)\Big],$$

Then from the definition of the determinant we have that

$$\tilde{\epsilon}_t = \det\big[\epsilon_t \;\; M\big] = \sum_{a=1}^{d} \xi_{a,t} \sum_{v} C_{a,v} \prod_{j\in v} \Big(\frac{1-b_i^{-1}L}{1-b_iL}\Big), \tag{A.16}$$

where the sum in v is over all combinations of indexes $\{1, 2, \cdots, q - r_\Theta\}$ with no repetition, so that $C_{a,v} = \det(M_{a,1}) \neq 0$ for some nonempty v and some a.

Now, Lemma 2 implies that there exists at least one a such that $\sum_v C_{a,v} \prod_{j\in v} \Big(\frac{1-b_i^{-1}L}{1-b_iL}\Big)\xi_{a,t}$ is nonlinearly predictable. Since the linear transformation can not change the nonlinear properties (in our case nonlinear predictability), $\tilde{\epsilon}_t$ and therefore ϵ_t is not a MDS.

References

Alessi, L., Barigozzi, M., and Capasso, M., 2011. Non-fundamentalness in structural econometric models: A review. *International Statistical Review*, 79(1):16–47.

Barro, R. J. and Redlick, C. J., 2011. Macroeconomic effects from government purchases and taxes. *The Quarterly Journal of Economics*, 126:51–102.

Blanchard, O. J. and Perotti, R., 2002. An empirical characterization of the dynamic effects of changes in government spending and taxes on output. *The Quarterly Journal of Economics*, 117(4):1329–1368.

Blanchard, O. J., L'Huillier, J.-P., and Lorenzoni, G., 2013. News, noise, and fluctuations: An empirical exploration. *American Economic Review*, 103(7):3045–70.

Box, G. E. and Pierce, D. A., 1970. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65(332):1509–1526.

Breidt, F. J. and Davis, R. A., 1992. Time-reversibility, identifiability and independence of innovations for stationary time series. *Journal of Time Series Analysis*, 13(5):377–390.

Brockwell, P. J. and Davis, R. A., 1991. *Time Series: Theory and Methods: Theory and Methods*. Springer series in statistics. Springer.

Chen, B., Choi, J., and Escanciano, J. C., 2012. Testing for fundamental vector moving average representations.

Domínguez, M. A. and Lobato, I. N., 2003. Testing the martingale difference hypothesis. *Econometric Reviews*, 22(4):351–377.

Escanciano, J. C. and Velasco, C., 2006. Generalized spectral tests for the martingale difference hypothesis. *Journal of Econometrics*, 134(1):151–185.

Fernández-Villaverde, J., Rubio-Ramírez, J. F., Sargent, T. J., and Watson, M. W., 2007. ABCs (and Ds) of Understanding VARs. *American Economic Review*, 97(3):1021–1026.

Forni, M. and Gambetti, L., 2014. Sufficient information in structural vars. *Journal of Monetary Economics*, 66:124–136.

Forni, M., Gambetti, L., and Sala, L., 2014. No news in business cycles. *The Economic Journal*.

Galí, J., López-Salido, J. D., and Vallés, J., 2007. Understanding the effects of government spending on consumption. *Journal of the European Economic Association*, 5(1):227–270.

Giannone, D. and Reichlin, L., 2006. Does information help recovering structural shocks from past observations? *Journal of the European Economic Association*, 4(2-3):455–465.

Harvey, C. R. and Siddique, A., 1999. Autoregressive conditional skewness. *Journal of financial and quantitative analysis*, 34(04):465–487.

Harvey, C. R. and Siddique, A., 2000. Conditional skewness in asset pricing tests. *The Journal of Finance*, 55(3):1263–1295.

Hong, Y., 1999. Hypothesis testing in time series via the empirical characteristic function: A generalized spectral density approach. *Journal of the American Statistical Association*, 94(448):1201–1220.

Hong, Y. and Lee, Y.-J., 2005. Generalized spectral tests for conditional mean models in time series with conditional heteroscedasticity of unknown form. *The Review of Economic Studies*, 72(2):499–541.

Jondeau, E. and Rockinger, M., 2003. Conditional volatility, skewness, and kurtosis: existence, persistence, and comovements. *Journal of Economic Dynamics and Control*, 27(10):1699–1737.

Leeper, E. M., Walker, T. B., and Yang, S.-C. S., 2013. Fiscal foresight and information flows. *Econometrica*, 81(3):1115–1145.

Lii, K.-S. and Rosenblatt, M., 1982. Deconvolution and estimation of transfer function phase and coefficients for nongaussian linear processes. *The Annals of Statistics*, pages 1195–1208.

Lippi, M. and Reichlin, L., 1994. Var analysis, nonfundamental representations, blaschke matrices. *Journal of Econometrics*, 63(1):307–325.

Ljung, G. M. and Box, G. E., 1978. On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303.

Mountford, A. and Uhlig, H., 2009. What are the effects of fiscal policy shocks? *Journal of applied econometrics*, 24(6):960–992.

Ramey, V. A., 2011. Identifying government spending shocks: it's all in the timing*. *The Quarterly Journal of Economics*, 126(1):1–50.

Ravenna, F., 2007. Vector autoregressions and reduced form representations of dsge models. *Journal of monetary economics*, 54(7):2048–2064.

Rosenblatt, M., 2000. *Gaussian and non-Gaussian linear time series and random fields.* Springer.

Rozanov, Y. A., 1967. *Stationary random processes.* Holden-Day.

Schmitt-Grohé, S. and Uribe, M., 2012. What's news in business cycles. *Econometrica*, 80(6):2733–2764.

Soccorsi, S., 2016. Measuring nonfundamentalness for structural vars. *ECARES Working Papers.*

News, Noise, and Tests of Asset Pricing Models
Supplementary Materials

Mehdi Hamidi Sahneh *
Universidad Carlos III de Madrid

I use the following notation throughout the Appendix. Let $\{\epsilon_t \equiv \epsilon_t(\theta_0)\}_{t=1}^{\infty}$ denotes the unobserved residuals and $\{\hat{\epsilon}_t\}_{t=1}^{\infty}$ the estimated residuals, which may contain some initial value. C denotes a generic bounded constant, $\|.\|$ the Euclidean norm, and A^* the complex conjugate of A. Let

$$\hat{f}^{(0,1,0)}(\omega, 0, v) \equiv \frac{1}{2\pi} \sum_{j=T-1}^{T-1} (1 - \frac{|j|}{T})^{1/2} k(j/h) \hat{\sigma}_j^{(1,0)}(0, v) e^{-ij\omega}, \quad \omega \in [-\pi, \pi]$$

where $\hat{\sigma}_j^{(1,0)}(0, v) = \frac{\partial}{\partial u} \hat{\sigma}_j(u, v)\big|_{u=0}$, $\hat{\sigma}_j(u, v) = \hat{\varphi}_j(u, v) - \hat{\varphi}_j(u, 0)\hat{\varphi}_j(0, v)$, and

$$\hat{\varphi}_j(u, v) = \frac{1}{T - |j|} \sum_{t=j+1}^{T} e^{iu'\hat{\epsilon}_t + iv'\hat{\epsilon}_{t-|j|}} \tag{C.1}$$

Throughout, I use

$$\sum_{j=1}^{T-1} a_T(j) = \sum_{j=1}^{T-1} k^2(j/h) T_j^{-1} = O_P(h/T) \tag{C.2}$$

given $h = cT^{\lambda}$ for $\lambda \in (0, \frac{1}{2})$, as shown in Hong (1999, (A.15), p.1213). The test statistic, that is robust to conditional heteroscedasticity and other time-varying higher order conditional moments of unknown form, is given as follows:

$$\hat{M} \equiv \left[\sum_{j=1}^{T-1} k^2(j/h) T_j \int \left\| \hat{\sigma}_j^{(1,0)}(0, v) \right\|^2 d\mathcal{W}(v) - \hat{C} \right] \Big/ \sqrt{\hat{D}} \tag{C.3}$$

where $T_j = T - j$, $\mathcal{W}(v) = \prod_{c=1}^{d} W(v_c)$, $W : \mathbb{R} \to \mathbb{R}^+$ is a nondecreasing function that weighs sets symmetric about zero equally, and the unspecified integrals are taken over the support of $\mathcal{W}(\cdot)$. $\hat{C}$ and $\hat{D}$ are estimate of the mean and the variance of $T \iint_{-\pi}^{\pi} \|\hat{f}^{(0,1,0)}(\omega, 0, v) -$

*Departamento de Economia, Universidad Carlos III de Madrid, Getafe, 28903, Spain. Email address: *mhamidis@eco.uc3m.es*.

$\hat{f}_0^{(0,1,0)}(\omega,0,v)\|^2 d\omega d\mathcal{W}(v)$,

$$\hat{C} = \sum_{j=1}^{T-1} k^2(j/h)T_j^{-1} \sum_{t=j+1}^{T-1} \|\hat{\epsilon}_t\|^2 \int \left|\hat{\psi}_{t-j}(v)\right|^2 d\mathcal{W}(v)$$

$$\hat{D} = 2\hat{s}^4 \sum_{j=1}^{T-2}\sum_{l=1}^{T-2} k^2(j/h)k^2(l/h) \iint \left|\hat{\sigma}_{j-l}(u,v)\right|^2 d\mathcal{W}(u)d\mathcal{W}(v)$$

where $\hat{\psi}_t(v) = e^{iv'\hat{\epsilon}_t} - T^{-1}\sum_{t=1}^T e^{iv'\hat{\epsilon}_t}$, and $\hat{s}^4 = \sum_{a,b=1}^d \left(T^{-1}\sum_{t=1}^T \hat{\epsilon}_{at}\hat{\epsilon}_{bt}\right)^2$.

Proof of Theorem 4.1(a): In order to prove $\hat{M} \xrightarrow{d} N(0,1)$, it suffices to show Theorems C1-C3 below. Theorem C1 implies that the use of $\{\hat{\epsilon}_t\}_{t=1}^T$ instead of $\{\epsilon_t\}_{t=1}^T$ has no impact on the limit distribution of $\hat{M}$. Theorem C2 implies the use of the truncated Wold innovations $\{\epsilon_{q,t}\}_{t=1}^T$ rather than $\{\epsilon_t\}_{t=1}^T$ has no impact on the limit distribution of $\hat{M}$ for q sufficiently large. This is used in Theorem C3 to simplify the proof of asymptotic normality of $\hat{M}$.

Theorem C1: Let M be defined in the same way as $\hat{M}$ in (C.3), with the unobservable $\{\epsilon_t\}_{t=1}^\infty$ replacing the estimated residual $\{\hat{\epsilon}_t\}_{t=1}^\infty$. Then under the conditions of Theorem 4.1(a), $\hat{M}-M \xrightarrow{p} 0$.

Theorem C2: Let M_q be defined as M with $\{\epsilon_{q,t}\}_{t=1}^T$ replacing $\{\epsilon_t\}_{t=1}^T$, where $\{\epsilon_{q,t}\}$ is as in Assumption A2. Then under the conditions of Theorem 4.1(a) and $q = h^{1+\frac{1}{4b-2}}(\ln^2 T)^{\frac{1}{2b-1}}$, $M_q - M \xrightarrow{p} 0$.

Theorem C3: Under the conditions of Theorem 4.1(a) and $q = h^{1+\frac{1}{4b-2}}(\ln^2 T)^{\frac{1}{2b-1}}$, $M_q \xrightarrow{p} N(0,1)$.

Proof of Theorem C1: Let $T_j \equiv T - |j|$, and let $\tilde{\sigma}_j^{(1,0)}(0,v)$ be defined in the same way as $\hat{\sigma}_j^{(1,0)}(0,v)$ in (C.1), with $\{\epsilon_t\}_{t=1}^T$ replacing $\{\hat{\epsilon}_t\}_{t=1}^T$. To show $\hat{M} - M \xrightarrow{p} 0$, it suffices to show

$$\hat{D}^{-\frac{1}{2}} \int \sum_{j=1}^{T-1} k^2(j/h)T_j\left[\|\hat{\sigma}_j^{(1,0)}(0,v)\|^2 - \|\tilde{\sigma}_j^{(1,0)}(0,v)\|^2\right] dW(v) \xrightarrow{p} 0 \tag{C.4}$$

$$\hat{C} - \tilde{C} = O_P(T^{-\frac{1}{2}}) \tag{C.5}$$

$$\hat{D} - \tilde{D} \xrightarrow{p} 0 \tag{C.6}$$

where $\tilde{C}$ and $\tilde{D}$ are defined the same way as $\hat{C}$ and $\hat{D}$, with $\{\epsilon_t\}_{t=1}^T$ replacing $\{\hat{\epsilon}_t\}_{t=1}^T$. For space, I focus on the proof of (C.4); the proofs for (C.5) and (C.6) are straightforward. To show (C.4), I first decompose

$$\int \sum_{j=1}^{T-1} k^2(j/h)T_j \Big[\|\hat{\sigma}_j^{(1,0)}(0,v)\|^2 - \|\tilde{\sigma}_j^{(1,0)}(0,v)\|^2\Big] dW(v) = \hat{A}_1 + 2\mathrm{Re}(\hat{A}_2) \tag{C.7}$$

where

$$\hat{A}_1 = \int \sum_{j=1}^{T-1} k^2(j/h)T_j \big\|\hat{\sigma}_j^{(1,0)}(0,v) - \tilde{\sigma}_j^{(1,0)}(0,v)\big\|^2 dW(v)$$

$$\hat{A}_2 = \int \sum_{j=1}^{T-1} k^2(j/h)T_j \Big\|\big[\hat{\sigma}_j^{(1,0)}(0,v) - \tilde{\sigma}_j^{(1,0)}(0,v)\big]\tilde{\sigma}_j^{(1,0)}(0,v)^*\Big\| dW(v)$$

where $\mathrm{Re}(\hat{A}_2)$ is the real part of $\hat{A}_2$ and $\tilde{\sigma}_j^{(1,0)}(0,v)^*$ is the complex conjugate of $\tilde{\sigma}_j^{(1,0)}(0,v)$. Then, (C.4) follows from the proofs of Propositions C1 and C2 bellow as $h \to \infty$ and $T \to \infty$.

Proposition C1: Under the conditions of Theorem 4.1(a), $\hat{A}_1 = O_P(1)$.

Proposition C2: Under the conditions of Theorem 4.1(a), $h^{-\frac{1}{2}}\hat{A}_2 \xrightarrow{p} 0$.

Proof of Proposition C1: Put $\hat{\delta}_t(v) \equiv e^{iv'\hat{\epsilon}_t} - e^{iv'\epsilon_t}$ and $\psi_t(v) \equiv e^{iv'\epsilon_t} - \mathrm{E}(e^{iv'\epsilon_t})$. Then straightforward algebra yields that for $j > 0$,

$$\begin{aligned}
\hat{\sigma}_j^{(1,0)}(0,v) - \tilde{\sigma}_j^{(1,0)}(0,v) &= iT_j^{-1} \sum_{t=j+1}^{T} (\hat{\epsilon}_t - \epsilon_t)\hat{\delta}_{t-j}(v) \qquad \text{(C.8)}\\
&- i\Big[T_j^{-1} \sum_{t=j+1}^{T} (\hat{\epsilon}_t - \epsilon_t)\Big]\Big[T_j^{-1} \sum_{t=j+1}^{T} \hat{\delta}_{t-j}(v)\Big]\\
&+ iT_j^{-1} \sum_{t=j+1}^{T} \epsilon_t \hat{\delta}_{t-j}(v)\\
&- i\Big(T_j^{-1} \sum_{t=j+1}^{T} \epsilon_t\Big)\Big[T_j^{-1} \sum_{t=j+1}^{T} \hat{\delta}_{t-j}(v)\Big]\\
&+ iT^{-1} \sum_{t=j+1}^{T} (\hat{\epsilon}_t - \epsilon_t)\psi_{t-j}(v)\\
&- i\Big[T_j^{-1} \sum_{t=j+1}^{T} (\hat{\epsilon}_t - \epsilon_t)\Big]\Big[T_j^{-1} \sum_{t=j+1}^{T} \psi_{t-j}(v)\Big]\\
&= i\Big[\hat{B}_{1j}(v) - \hat{B}_{2j}(v) + \hat{B}_{3j}(v) - \hat{B}_{4j}(v) + \hat{B}_{5j}(v) - \hat{B}_{6j}(v)\Big], \qquad \text{(C.9)}
\end{aligned}$$

Say. It follows that $\hat{A}_1 \le 8\sum_{a=1}^{6}\sum_{j=1}^{T} k^2(j/h)T_j \int \|\hat{B}_{aj}(v)\|^2 dW(v)$. Proposition C1 follows from Lemmas C1 to C6 below, and $h/T \to 0$.

Lemma C1: $\sum_{j=1}^{T} k^2(j/h)T_j \int \|\hat{B}_{1j}(v)\|^2 dW(v) = O_P(h/T)$.

Lemma C2: $\sum_{j=1}^{T} k^2(j/h)T_j \int \|\hat{B}_{2j}(v)\|^2 dW(v) = O_P(h/T)$.

Lemma C3: $\sum_{j=1}^{T} k^2(j/h)T_j \int \|\hat{B}_{3j}(v)\|^2 dW(v) = O_P(h/T)$.

Lemma C4: $\sum_{j=1}^{T} k^2(j/h)T_j \int \|\hat{B}_{4j}(v)\|^2 dW(v) = O_P(h/T)$.

Lemma C5: $\sum_{j=1}^{T} k^2(j/h)T_j \int \|\hat{B}_{5j}(v)\|^2 dW(v) = O_P(1)$.

Lemma C6: $\sum_{j=1}^{T} k^2(j/h)T_j \int \|\hat{B}_{6j}(v)\|^2 dW(v) = O_P(h/T)$.

Proof of Lemma C1: By the Cauchy-Schwarz inequality and the inequality that $|e^{iz_1} - e^{iz_2}| \leq |z_1 - z_2|$ for any real-valued variables z_1 and z_2, we have

$$\begin{aligned} \|\hat{B}_{1j}(v)\|^2 &\leq \Big[T_j^{-1} \sum_{t=1}^{T} \|\hat{\epsilon}_t - \epsilon_t\|^2\Big] \Big[T_j^{-1} \sum_{t=1}^{T} |\hat{\delta}_t(v)|^2\Big] \\ &\leq \ \|v\|^2 \Big[T_j^{-1} \sum_{t=1}^{T} \|\hat{\epsilon}_t - \epsilon_t\|^2\Big]^2. \end{aligned} \tag{C.10}$$

It follows from Assumptions A4-A6 that

$$\begin{aligned} &\int \sum_{j=1}^{T-1} k^2(j/h)T_j \|\hat{B}_{1j}(v)\|^2 dW \\ &\qquad \leq \Big[\sum_{j=1}^{T-1} k^2(j/h)T_j^{-1}\Big] \Big[\sum_{t=1}^{T} \|\hat{\epsilon}_t - \epsilon_t\|^2\Big]^2 \times \int \|v\|^2 dW(v) = O_P(h/T) \end{aligned} \tag{C.11}$$

where I have used the fact that

$$\sum_{t=1}^{T} \|\hat{\epsilon}_t - \epsilon_t\|^2 = O_P(1), \tag{C.12}$$

which has been shown in (Ling and McAleer, 2003, (B.18), p.302). □

Proof of Lemma C2: By the inequality that $|e^{iz_1} - e^{iz_2}| \leq |z_1 - z_2|$ for any real-valued variables z_1 and z_2, we have

$$\begin{aligned} \|\hat{B}_{2j}(v)\|^2 &\leq \Big[T_j^{-1} \sum_{t=1}^{T} \|\hat{\epsilon}_t - \epsilon_t\|^2\Big] \Big[T_j^{-1} \sum_{t=1}^{T} \|v'\hat{\epsilon}_t - v'\epsilon_t\|^2\Big] \\ &\leq \|v\|^2 \Big[T_j^{-1} \sum_{t=1}^{T} \|\hat{\epsilon}_t - \epsilon_t\|^2\Big]^2 \end{aligned}$$

The rest of the proof is similar to that of Lemma C1. □

Proof of Lemma C3: I decompose

$$\hat{B}_{3j}(v) = T_j^{-1} \sum_{t=j+1}^{T} \epsilon_t \Big[e^{iv'\hat{\epsilon}_{t-j}} - e^{iv'\epsilon_{t-j}(\hat{\theta})}\Big] + T_j^{-1} \sum_{t=j+1}^{T} \epsilon_t \Big[e^{iv'\epsilon_{t-j}(\hat{\theta})} - e^{iv'\epsilon_{t-j}}\Big]$$
$$\equiv \hat{B}_{31j}(v) + \hat{B}_{32j}(v). \tag{C.13}$$

First, I consider $\hat{B}_{31j}(v)$. By the Cauchy-Schwarz inequality, Minkowski's inequality, and $|e^{iz_1} - e^{iz_2}| \leq |z_1 - z_2|$ we have

$$\begin{aligned}
\mathrm{E}\|\hat{B}_{31j}(v)\| &\leq \|v\|^2 \mathrm{E}\Bigg[T_j^{-1} \sum_{t=j+1}^{T} \|\epsilon_t\| \|\hat{\epsilon}_{t-j} - \epsilon_{t-j}(\hat{\theta})\|\Bigg]^2 \\
&\leq \|v\|^2 \Bigg\{T_j^{-1} \sum_{t=j+1}^{T} \Big(\mathrm{E}\|\epsilon_t^4\|\Big)^{1/4} \Big[\mathrm{E}\big(\sup_{\theta\in\Theta} \|\epsilon_{t-j}(\hat{I}_{t-1}, \theta) - \epsilon_{t-j}(I_{t-1}, \theta)\|\big)^4\Big]^{1/4}\Bigg\}^2 \\
&\leq C T_j^{-2} \|v\|^2
\end{aligned}$$

It follows from Markov's inequality, (C.2), and Assumptions A1, A4 and A6 that

$$\sum_{j=1}^{T-1} k^2(j/h) T_j \|\hat{B}_{31j}(v)\|^2 = O_P(h/T). \tag{C.14}$$

Next, consider $\hat{B}_{32j}(v)$. Using the inequality that $|e^{iz} - 1 - iz| \leq |z|^2$ for any real-valued z, and a second order Taylor series expansion, we have

$$\begin{aligned}
T_j \|\hat{B}_{32j}(v)\| \leq \|v'\| \|\hat{\theta} - \theta_0\| &\sum_{t=j+1}^{T} \|\epsilon_t\| \Bigg\| \frac{\partial \epsilon_{t-j}(\theta_0)}{\partial \theta} e^{iv'\epsilon_{t-j}} \Bigg\| \\
&+ \|v'\|^2 \sum_{t=j+1}^{T} \|\epsilon_t\| \Big\| \epsilon_{t-j}(\hat{\theta}) - \epsilon_{t-j} \Big\|^2 \\
&+ \|v'\| \|\hat{\theta} - \theta_0\|^2 \sum_{t=j+1}^{T} \|\epsilon_t\| \sup_{\theta\in\Theta} \Bigg\| \frac{\partial^2 \epsilon_{t-j}(\theta_0)}{\partial \theta \partial \theta'} \Bigg\|
\end{aligned}$$

It follows from Assumptions A1-A7 and (C.2) that

$$
\begin{aligned}
\sum_{j=1}^{T-1} \int k^2(j/h) T_j \|\hat{B}_{32j}(v)\|^2 dW(v) \leq 4\|\sqrt{T}(\hat{\theta}-\theta_0)\|^2 \sum_{j=1}^{T-1} k^2(j/h) \\
\times \int \Big\| T_j^{-1} \sum_{t=j+1}^{T} \|\epsilon_t\|^2 \Big\| \frac{\partial \epsilon_{t-j}(\theta_0)}{\partial \theta} e^{iv'\epsilon_{t-j}} \Big\|^2 \|v\|^2 dW(v) \\
+ 4\|\hat{\theta}-\theta\|^4 \sum_{j=1}^{T-1} a_T(j) \int \|v'\|^4 dW(v) \sum_{t=1}^{T} \|\epsilon_t\|^2 \sum_{t=1}^{T} \sup_{\theta\in\Theta} \Big\| \frac{\partial \epsilon_t(\theta_0)}{\partial \theta} \Big\|^4 \\
+ 4\|\hat{\theta}-\theta\|^4 \sum_{j=1}^{T-1} a_T(j) \int \|v'\|^2 dW(v) \sum_{t=1}^{T} \|\epsilon_t\|^2 \sum_{t=1}^{T} \sup_{\theta\in\Theta} \Big\| \frac{\partial^2 \epsilon_t(\theta)}{\partial \theta \partial \theta'} \Big\|^2 \\
= O_P(h/T) \qquad \text{(C.15)}
\end{aligned}
$$

where $\bar{\theta}$ is between $\hat{\theta}$ and θ_0. The desired results follows from $\mathrm{E}\Big\| \sum_{t=j+1}^{T} \epsilon_t \frac{\partial \epsilon_{t-j}(\theta_0)}{\partial \theta} e^{iv'\epsilon_{t-j}} \Big\|^2 \leq CT_j$ given $\mathrm{E}(\epsilon_t|I_{t-1}) = 0$ $a.s.$ under $\mathbb{H}_0$, Assumption A1, and the fact that

$$
\frac{1}{T} \sum_{t=1}^{T} \sup_{\theta\in\Theta} \Big\| \frac{\partial^2 \epsilon_{tl}(\theta)}{\partial \theta \partial \theta'} \Big\|^2 \leq \frac{1}{T} \sum_{t=1}^{T} \Big(C + \sum_{k=1}^{\infty} \rho^k \|x_{t-k}\|^2 \Big) = O_P(1) \qquad \text{(C.16)}
$$

where $0 < \rho < 1$ and $l = 1, \cdots, d$. □

Proof of Lemma C4: By the Cauchy-Schwarz inequality, and $|\hat{\delta}_t(v)| \leq \|v\|.\|\hat{\epsilon}_t - \epsilon_t\|$ we have

$$
\begin{aligned}
\sum_{j=1}^{T-1} k^2(j/h) T_j \int \|\hat{B}_{4j}(v)\|^2 dW(v) \\
\leq \sum_{j=1}^{T-1} k^2(j/h) \Big(T_j^{-1} \sum_{t=j+1}^{T} \|\epsilon_t\| \Big)^2 \Big[\sum_{t=1}^{T} \|\hat{\epsilon}_t - \epsilon_t\|^2 \Big] \int \|v\|^2 dW(v) \\
= O_P(h/T)
\end{aligned}
$$

given (C.2) and the fact that $\mathrm{E}(T_j^{-1} \sum_{t=j+1}^{T} \epsilon_t)^2 = O(1)$ by the MDS property of $\{\epsilon_t\}$ under $\mathbb{H}_0$. □

Proof of Lemma C5: First decompose

$$
\begin{aligned}
\hat{B}_{5j}(v) = T_j^{-1} \sum_{t=j+1}^{T} [\hat{\epsilon}_t - \epsilon_t(\hat{\theta})]\psi_{t-j}(v) + T_j^{-1} \sum_{t=j+1}^{T} [\epsilon_t(\hat{\theta}) - \epsilon_t]\psi_{t-j}(v) \\
\equiv \hat{B}_{51j}(v) + \hat{B}_{52j}(v). \qquad \text{(C.17)}
\end{aligned}
$$

Given $|\psi_t(v)| \leq 2$, (C.2), and Assumptions A4 and A5, we have

$$\sum_{j=1}^{T-1} k^2(j/h)T_j \int \|\hat{B}_{51j}(v)\|^2 dW(v) \leq 4\Big[\sum_{t=1}^{T} \|\hat{\epsilon}_t - \epsilon_t(\hat{\theta})\|\Big]^2 \times \sum_{j=1}^{T-1} a_T(j) \int dW(v) = O_P(h/T) \tag{C.18}$$

Also, by a second order Taylor series expansion, we have

$$\begin{aligned}
&\sum_{j=1}^{T-1} k^2(j/h)T_j \int \|\hat{B}_{52j}(v)\|^2 dW(v) \\
&\leq 2\|\sqrt{T}(\hat{\theta} - \theta_0)\|^2 \sum_{j=1}^{T-1} k^2(j/h) \int \Big\|T_j^{-1} \sum_{t=j+1}^{T} \frac{\partial \epsilon_t(\theta_0)}{\partial \theta} \psi_{t-j}(v)\Big\|^2 dW(v) \\
&+ 2\|\sqrt{T}(\hat{\theta} - \theta_0)\|^4 \Big[T^{-1} \sum_{t=1}^{T} \sup_{\theta \in \Theta} \Big\|\frac{\partial^2 \epsilon_t(\theta)}{\partial \theta \partial \theta'}\Big\|\Big]^2 \sum_{j=1}^{T-1} a_T(j) \int dW(v) \\
&= O_P(1) + O_P(h/T)
\end{aligned} \tag{C.19}$$

where the first terms is $O_P(h/T)$ given (C.2), and the last term is $O_P(1)$, as is shown in Hong and Lee (2005, (A.16), p.530).□

Proof of Lemma C6: The proof is similar to the proof of Lemma C4 and therefore omitted.□

Proof of Proposition C2: Given the decomposition in (C.8), we have

$$\Big\|[\hat{\sigma}_j^{(1,0)}(0, v) - \tilde{\sigma}_j^{(1,0)}(0, v)]\tilde{\sigma}_j^{(1,0)}(0, v)^*\Big\| \leq \sum_{a=1}^{6} \|\hat{B}_{aj}(v)\| \|\tilde{\sigma}_j^{(1,0)}(0, v)\| \tag{C.20}$$

where $\hat{B}_{aj}(v)$ are defined in (C.8). First consider the cases $a = 1, 2, 3, 4$, and 6. By the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
&\sum_{j=1}^{T-1} k^2(j/h)T_j \int \|\hat{B}_{aj}(v)\|^2 \|\tilde{\sigma}_j^{(1,0)}(0, v)\| dW(v) \\
&\leq \Big[\sum_{j=1}^{T-1} k^2(j/h)T_j \int \|\hat{B}_{aj}(v)\|^2 dW(v)\Big]^{\frac{1}{2}} \Big[\sum_{t=1}^{T-1} k^2(j/h)T_j \int \|\tilde{\sigma}_j^{(1,0)}(0, v)\|^2 dW(v)\Big]^{\frac{1}{2}} \\
&= O_P(h^{\frac{1}{2}}/T^{\frac{1}{2}})O_P(h^{\frac{1}{2}}) = o_P(h^{\frac{1}{2}}), \quad a = 1, 2, 3, 4, 6,
\end{aligned}$$

given Lemmas C1-C4 and C6, and $h/T \to 0$, using the fact that

$$h^{-1}\sum_{j=1}^{T-1} k^2(j/h)T_j \int \|\tilde{\sigma}_j^{(1,0)}(0,v)\|^2 dW(v) = O_P(1)$$

by Markov's inequality, the MDS property of $\{\epsilon_t\}$ and (C.2).

It remains to consider the case $a = 5$. By (C.17) and the triangle inequality, we have

$$\begin{aligned}
&\sum_{j=1}^{T-1} k^2(j/h)T_j \int \|\hat{B}_{5j}(v)\|\|\tilde{\sigma}_j^{(1,0)}(0,v)\| dW(v) \\
&\qquad \le \sum_{j=1}^{T-1} k^2(j/h)T_j \int \|\hat{B}_{51j}(v)\|\|\tilde{\sigma}_j^{(1,0)}(0,v)\| dW(v) \\
&\qquad + \sum_{j=1}^{T-1} k^2(j/h)T_j \int \|\hat{B}_{52j}(v)\|\|\tilde{\sigma}_j^{(1,0)}(0,v)\| dW(v). \qquad (C.21)
\end{aligned}$$

For the first term in (C.21), we have

$$\begin{aligned}
&\sum_{j=1}^{T-1} k^2(j/h)T_j \int \|\hat{B}_{51j}(v)\|\|\tilde{\sigma}_j^{(1,0)}(0,v)\| dW(v) \\
&\qquad \le 2\Big[\sum_{t=j+1}^{T} \|\hat{\epsilon}_t - \epsilon_t(\hat{\theta})\|\Big] \times \Big[\sum_{j=1}^{T-1} k^2(j/h) \int \|\tilde{\sigma}_j^{(1,0)}(0,v)\|\Big] dW(v) \\
&\qquad = O_P(h/T^{\frac{1}{2}}) \qquad (C.22)
\end{aligned}$$

given (C.2) and the MDS property of $\{\epsilon_t\}$. For the second term in (C.21) we have

$$\begin{aligned}
&\sum_{j=1}^{T-1} k^2(j/h)T_j \int \|\hat{B}_{52j}(v)\|\|\tilde{\sigma}_j^{(1,0)}(0,v)\| dW(v) \\
&\le \|\hat{\theta}-\theta_0\| \sum_{j=1}^{T-1} k^2(j/h)T_j \int \Big\| T_j^{-1} \sum_{t=j+1}^{T} \frac{\partial \epsilon_t(\theta_0)}{\partial \theta} \psi_{t-j}(v) \Big\| \|\tilde{\sigma}_j^{(1,0)}(0,v)\| dW(v) \\
&+ \|\sqrt{T}(\hat{\theta}-\theta_0)\|^2 \Big[T^{-1}\sum_{t=1}^{T} \sup_{\theta\in\Theta} \Big\|\frac{\partial^2 \epsilon_t(\theta)}{\partial\theta\partial\theta'}\Big\|\Big] \sum_{j=1}^{T-1} k^2(j/h) \int \|\tilde{\sigma}_j^{(1,0)}(0,v)\| dW(v) \\
&= O_P(1 + h/T^{\frac{1}{2}}) + O_P(h/T^{\frac{1}{2}}) = o_P(h^{\frac{1}{2}}) \qquad (C.23)
\end{aligned}$$

given $p \to \infty$, $h/T \to 0$, and Assumptions A3-A7, using $T_j \mathrm{E}\|\tilde{\sigma}_j^{(1,0)}(0,v)\|^2 \le C$ under the MDS property of $\{\epsilon_t\}$.$\square$

Proof of Theorem C2: Let $\hat{A}_{1q}$ and $\hat{A}_{2q}$ be defined in the same way as $\hat{A}_1$ and $\hat{A}_2$ in (C.7), with $\{\epsilon_{q,t}\}_{t=1}^{T}$ replacing $\{\hat{\epsilon}_t\}_{t=1}^{T}$. It suffices to show $h^{-\frac{1}{2}}\hat{A}_{1q} \xrightarrow{p} 0$ and $h^{-\frac{1}{2}}\hat{A}_{2q} \xrightarrow{p} 0$. Put

$\delta_{q,t} \equiv e^{iv'\epsilon_t} - e^{iv'\epsilon_{q,t}}$ and $\psi_{q,t}(v) \equiv e^{iv'\epsilon_{q,t}} - \mathrm{E}(e^{iv'\epsilon_{q,t}})$. Let $\tilde{\sigma}_{q,j}^{(1,0)}(0,v)$ be defined as $\tilde{\sigma}_j^{(1,0)}(0,v)$, with $\{\epsilon_{q,t}\}_{t=1}^T$ replacing $\{\hat{\epsilon}_t\}_{t=1}^T$. Then, similar to (C.8), we have

$$\begin{aligned}
\hat{\sigma}_j^{(1,0)}(0,v) - \tilde{\sigma}_{q,j}^{(1,0)}(0,v) = {} & iT_j^{-1} \sum_{t=j+1}^{T} (\epsilon_t - \epsilon_{q,t})\delta_{q,t-j}(v) - i\Big[T_j^{-1}\sigma_{t=j+1}^{T}(\epsilon_t - \epsilon_{q,t})\Big] \\
& \times \Big[T_j^{-1}\sum_{t=j+1}^{T}\delta_{q,t-j}(v)\Big] + iT_j^{-1}\sum_{t=j+1}^{T}\epsilon_{q,t}\delta_{q,t-j}(v) \\
& - i\Big(T_j^{-1}\sum_{t=j+1}^{T}\epsilon_{q,t}\Big)\Big[T_j^{-1}\sum_{t=j+1}^{T}\delta_{q,t-j}(v)\Big] \\
& + iT^{-1}\sum_{t=j+1}^{T}(\epsilon_t - \epsilon_{q,t})\psi_{q,t-j}(v) \\
& - i\Big[T_j^{-1}\sum_{t=j+1}^{T}(\epsilon_t - \epsilon_{q,t})\Big]\Big[T_j^{-1}\sum_{t=j+1}^{T}\psi_{q,t-j}(v)\Big] \\
& = i\Big[\hat{B}_{1jq}(v) - \hat{B}_{2jq}(v) + \hat{B}_{3jq}(v) - \hat{B}_{4jq}(v) + \hat{B}_{5jq}(v) - \hat{B}_{6jq}(v)\Big]
\end{aligned}$$

The proof is analogous to that of Theorem C1 noting that $\mathrm{E}(\epsilon_t|I_{t-1}) = 0$ a.s. and $\mathrm{E}(\epsilon_{q,t}|I_{t-1}) = 0$ a.s., and therefore is omitted. $\square$

Proof of Theorem C3: It follows from Propositions C3 and C4 below.

Proposition C3: Let $\tilde{\sigma}_{q,j}^{(1,0)}(0,v)$ and $\tilde{C}_q$ be defined as $\tilde{\sigma}_j^{(1,0)}(0,v)$ and $\tilde{C}$ respectively, with $\{\epsilon_{q,t}\}_{t=1}^T$ replacing $\{\epsilon_t\}_{t=1}^T$. Then, under the conditions of Theorem 4.1(a),

$$h^{-\frac{1}{2}}\sum_{j=1}^{T-1}k^2(j/h)T_j\int\|\tilde{\sigma}_{q,j}^{(1,0)}(0,v)\|^2 dW(v) = h^{-\frac{1}{2}}\tilde{C}_q + h^{-\frac{1}{2}}\tilde{V}_q + o_P(1),$$

where $\tilde{V}_q = \sum_{a=1}^d \sum_{t=2q+2}^T \epsilon_{aq,t}\sum_{j=1}^q a_T(j)\int \psi_{q,t-j}(v)\sum_{s=1}^{t-2q-1}\epsilon_{aq,s}\psi_{q,s-j}(v)^* dW(v)$.

Proposition C4: Let $\tilde{D}_q$ be defined as $\tilde{D}$ with $\{\epsilon_{q,t}\}_{t=1}^T$ replacing $\{\epsilon_t\}_{t=1}^T$. Then $\tilde{D}^{-\frac{1}{2}}\tilde{V}_q \xrightarrow{d} N(0,1)$.

Proof of Proposition C3: Recall that $\phi_q(v) = \mathrm{E}(e^{iv'\epsilon_{q,t}})$, $\psi_{q,t}(v) = e^{iv'\epsilon_{q,t}} - \phi_q(v)$, and

$\tilde{\sigma}_{q,j}^{(1,0)}(0,v) = T_j^{-1} \sum_{t=j+1}^{T} \epsilon_{q,t}\psi_{q,t}(v)$. First decompose

$$\begin{aligned}
\sum_{j=1}^{T-1} k^2(j/h)T_j \int \|\tilde{\sigma}_{q,j}^{(1,0)}(0,v)\|^2 dW(v) &= \sum_{j=1}^{T-1} a_T(j) \int \Big\| \sum_{t=1}^{T} \epsilon_{q,t}\psi_{q,t-j}(v) \Big\|^2 dW(v) \\
&+ \sum_{j=1}^{T-1} a_T(j) \int \Big\| \sum_{t=1}^{j} \epsilon_{q,t}\psi_{q,t-j}(v) \Big\|^2 dW(v) \\
&- 2\mathrm{Re} \sum_{j=1}^{T-1} a_T(j) \int \sum_{a=1}^{d} \big[\sum_{t=1}^{T} \epsilon_{aq,t}\psi_{q,t-j}(v) \big] \\
&\times \big[\sum_{t=1}^{j} \epsilon_{aq,t}\psi_{q,t-j}(v) \big]^* dW(v) \\
&\equiv \tilde{Q}_q + \tilde{R}_{1q} - 2\tilde{R}_{2q}, \qquad \text{(C.24)}
\end{aligned}$$

say, where $\epsilon_{aq,t}$ is the a-th component of $\epsilon_{q,t}$. Next write

$$\begin{aligned}
\tilde{Q}_q &= \sum_{j=1}^{T-1} a_T(j) \int \sum_{t=1}^{T} \|\epsilon_{q,t}\|^2 |\psi_{q,t-j}(v)|^2 dW(v) \\
&+ 2\mathrm{Re} \sum_{j=1}^{T-1} a_T(j) \int \sum_{a=1}^{d} \big[\sum_{t=2}^{T} \epsilon_{aq,t}\psi_{q,t-j}(v) \big] \big[\sum_{s=1}^{t-1} \epsilon_{aq,s}\psi_{q,t-j}(v) \big]^* dW(v) \\
&\equiv \tilde{C}_q + 2\mathrm{Re}\tilde{U}_q, \qquad \text{(C.25)}
\end{aligned}$$

where we further decompose

$$\begin{aligned}
\tilde{U}_q &= \sum_{a=1}^{d} \sum_{t=2q+2}^{T} \epsilon_{aq,t} \int \sum_{j=1}^{T-2} a_T(j)\psi_{q,t-j}(v) \sum_{s=1}^{t-2q-1} \epsilon_{aq,s}\psi_{q,s-j}(v)^* dW(v) \\
&+ \sum_{a=1}^{d} \sum_{t=2}^{T} \epsilon_{aq,t} \int \sum_{j=1}^{T-2} a_T(j)\psi_{q,t-j}(v) \sum_{s=\max(1,t-2q)}^{t-1} \epsilon_{aq,s}\psi_{q,s-j}(v)^* dW(v) \\
&\equiv \tilde{U}_{1q} + \tilde{R}_{3q}, \qquad \text{(C.26)}
\end{aligned}$$

where in the first term we have $t - s > 2q$ so that $\{\epsilon_{aq,t}, \psi_{q,t-j}(v)\}_{j=1}^{q}$ is independent of $\{\epsilon_{aq,s}, \psi_{q,s-j}(v)\}_{j=1}^{q}$ for q sufficiently large. In the second term $\tilde{R}_{3q}$, we have $0 < t - s < 2q$.

Finally, we have

$$\begin{aligned}\tilde{U}_{1q} &= \sum_{a=1}^{d}\sum_{t=2q+2}^{T}\epsilon_{aq,t}\sum_{j=1}^{q}a_T(j)\int\psi_{q,t-j}(v)\sum_{s=1}^{t-2q-1}\epsilon_{aq,s}\psi_{q,s-j}(v)^*dW(v)\\ &+\sum_{a=1}^{d}\sum_{t=2q+2}^{T}\epsilon_{aq,t}\sum_{j=q+1}^{T-1}a_T(j)\int\psi_{q,t-j}(v)\sum_{s=1}^{t-2q-1}\epsilon_{aq,s}\psi_{q,s-j}(v)^*dW(v)\\ &\equiv \tilde{V}_q+\tilde{R}_{4q}, \end{aligned} \tag{C.27}$$

where the first term is contributed by the lag orders j from 1 to q; and the second term is contributed by the lag orders $j > q$. It then follows from (C.24) – (C.27) that

$$\begin{aligned}\sum_{j=1}^{T-1}k^2(j/h)T_j\int\|\tilde{\sigma}_{q,j}^{(1,0)}(0,v)\|^2dW(v) &= \tilde{C}_q+2\mathrm{Re}\tilde{V}_q\\ &\quad+\tilde{R}_{1q}-2\mathrm{Re}\big(\tilde{R}_{2q}-\tilde{R}_{3q}-\tilde{R}_{4q}\big).\end{aligned}$$

It suffices to show Lemmas C7-C11 below, which imply $h^{-\frac{1}{2}}[\tilde{C}_q-\tilde{C}_q]=o_P(1)$ and $h^{-\frac{1}{2}}\tilde{R}_{aq}=o_P(1)$ for $a=1,2,3,4$ given $q=h^{1+\frac{1}{4b-2}}(\ln^2 T)^{\frac{1}{2b-1}}$ and $p=cT^{\lambda}$ for $0<\lambda<(3+\frac{1}{4b-2})^{-1}$.

Lemma C7: Let $\tilde{C}_q$ be defined as in (C.25). Then $\tilde{C}_q-\tilde{C}_q=O_P(h^2/T)$.

Lemma C8: Let $\tilde{R}_{1q}$ be defined as in (C.24). Then $\tilde{R}_{1q}=O_P(h^2/T)$.

Lemma C9: Let $\tilde{R}_{2q}$ be defined as in (C.24). Then $\tilde{R}_{2q}=O_P(h^{\frac{3}{2}}/T^{\frac{1}{2}})$.

Lemma C10: Let $\tilde{R}_{3q}$ be defined as in (C.26). Then $\tilde{R}_{3q}=O_P(q^{\frac{1}{2}}h/T^{\frac{1}{2}})$.

Lemma C11: Let $\tilde{R}_{4q}$ be defined as in (C.27). Then $\tilde{R}_{4q}=O_P(h^{2b}\ln(T)/q^{2b-1})$.

Proof of Lemma C7: By Markov's inequality and $\mathrm{E}|\tilde{C}_q-\tilde{C}_q|\le\frac{Ch^2}{T}$ given $\sum_{j=1}^{T-1}(j/p)a_T(j)=O(h/T)$.□

Proof of Lemma C8: By the MDS property of $\{\epsilon_{q,t},\mathcal{F}_{t-1}\}$, where $\mathcal{F}_{t-1}$ is the sigma-field generated by $\{\epsilon_{t-j}\}_{j=1}^{\infty}$, we can obtain

$$\mathrm{E}\int\|\sum_{t=1}^{j}\epsilon_{q,t}\psi_{q,t-j}(v)\|^2dW(v)=\sum_{t=1}^{j}\int\mathrm{E}\|\epsilon_{q,t}\|^2|\psi_{q,t-j}(v)|^2\le Cj.$$

The result then follows from Markov's inequality and $\sum_{j=1}^{T-1}(j/p)a_T(j)=O(h/T)$.□

Proof of Lemma C9: The proof is similar to Lemma C8, given the Assumption A5, and

$$\mathrm{E}\Big|\int\sum_{a=1}^{d}\sum_{t=1}^{T}\big[\sum_{t=1}^{j}\epsilon_{aq,t}\psi_{q,t-j}(v)\big]\big[\epsilon_{aq,t}\psi_{q,t-j}(v)\big]^*dW(v)\Big|\le C(jT)^{\frac{1}{2}}.$$

$\square$

Proof of Lemma C10: By the MDS property of $\{\epsilon_{q,t}, \mathcal{F}_{t-1}\}$, Minkowski's inequality and (C.2), we have

$$\begin{aligned}
\mathrm{E}\big|\tilde{R}_{3q}\big|^2 &= \sum_{a=1}^{d}\sum_{t=2}^{T}\mathrm{E}\Big|\sum_{j=1}^{T-1} a_T(j)\int \epsilon_{aq,t}\psi_{q,t-j}(v)\sum_{s=\max(1,t-2q)}^{t-1}\epsilon_{aq,s}\psi_{q,s-j}(v)^* dW(v)\Big|^2 \\
&\leq \sum_{a=1}^{d}\sum_{t=2}^{T}\left[\sum_{j=1}^{T-1} a_T(j)\int \left(\mathrm{E}\big|\epsilon_{aq,t}\psi_{q,t-j}(v)\sum_{s=\max(1,t-2q)}^{t-1}\epsilon_{aq,s}\psi_{q,s-j}(v)^*\big|^2\right)^{\frac{1}{2}} dW(v)\right]^2 \\
&\leq 2CTq\left[\sum_{j=1}^{T-1} a_T(j)\right]^2 = O(qh^2/T).\square
\end{aligned}$$

Proof of Lemma C11: By the MDS property of $\{\epsilon_{q,t}, \mathcal{F}_{t-1}\}$ and Minkowski's inequality, we have

$$\begin{aligned}
\mathrm{E}\big|\tilde{R}_{4q}\big|^2 &= \sum_{a=1}^{d}\sum_{t=2q+2}^{T}\mathrm{E}\Big|\sum_{j=q+1}^{T-1} a_T(j)\int \epsilon_{aq,t}\psi_{q,t-j}(v)\sum_{s=1}^{t-2q-1}\epsilon_{aq,s}\psi_{q,s-j}(v)^* dW(v)\Big|^2 \\
&\leq \sum_{a=1}^{d}\sum_{t=2q+2}^{T}\left[\sum_{j=q+1}^{T-1} a_T(j)\int \left(\mathrm{E}\Big|\epsilon_{aq,t}\psi_{q,t-j}(v)\sum_{s=1}^{t-2q-1}\epsilon_{aq,s}\psi_{q,s-j}(v)^*\Big|^2\right)^{\frac{1}{2}} dW(v)\right]^2 \\
&\leq CT^2\left[\sum_{j=q+1}^{T-1} a_T(j)\right]^2 \leq C^3T^2\left[\sum_{j=q+1}^{T-1}(j/p)^{-2b}T_j^{-1}\right]^2 \\
&= O(h^{4b}\ln^2(T)/q^{4b-2}).\square
\end{aligned}$$

Proof of Proposition C4: Let $\tilde{V}_q = \sum_{t=2q+2}^{T} V_q(t)$, where

$$V_q(t) = \sum_{a=1}^{d}\sum_{j=1}^{q} a_T(j)\int [\epsilon_{aq,t}\psi_{q,t-j}(v)]H_{j,t-2q-1}(v)dW(v),$$

and $H_{j,t-2q-1}(v) = \sum_{s=1}^{t-2q-1}[\epsilon_{aq,s}\psi_{q,s-j}(v)]^*$. The martingale limit theorem of Brown (1971) states that $\mathrm{Var}(2\mathrm{Re}\tilde{V}_q)^{-\frac{1}{2}} 2\mathrm{Re}\tilde{V}_q \xrightarrow{d} N(0,1)$ if

$$\mathrm{Var}(2\mathrm{Re}\tilde{V}_q)^{-1} \sum_{t=1}^{T} \big[2\mathrm{Re}V_q(t)\big]^2 \times \mathbf{1}\big[\mathrm{Var}(2\mathrm{Re}\tilde{V}_q)^{-\frac{1}{2}} |2\mathrm{Re}V_q(t)| > \eta\big] \to 0, \quad \forall \eta > 0 \tag{C.28}$$

$$\mathrm{Var}(2\mathrm{Re}\tilde{V}_q)^{-1} \sum_{t=1}^{T} \mathrm{E}\big[2\mathrm{Re}V_q^2(t)|\mathcal{F}_{t-1}\big] \xrightarrow{p} 1. \tag{C.29}$$

To apply this theorem, first I compute $\mathrm{Var}(2\mathrm{Re}\tilde{V}_q)$. Because $\mathrm{Var}(2\mathrm{Re}\tilde{V}_q) = \mathrm{E}(\tilde{V}_q^2) + \mathrm{E}(\tilde{V}_q^*)^2 + 2\mathrm{E}|\tilde{V}_q|^2$, and $\mathrm{E}(\tilde{V}_q^2) = \mathrm{E}(\tilde{V}_q^*)^2 = \mathrm{E}|\tilde{V}_q|^2$, we only need $\mathrm{E}(\tilde{V}_q^2)$. By the MDS property of $\{\epsilon_{q,t}, \mathcal{F}_{t-1}\}$ under $\mathbb{H}_0$ and independence of $\epsilon_{q,t}$ and $\{\epsilon_{t-j-1}\}_{j=q}^{\infty}$ for q sufficiently large, we have

$$\begin{aligned}
\mathrm{E}(\tilde{V}_q^2) &= \sum_{a,b=1}^{d} \sum_{t=2q+2}^{T} \mathrm{E}\Big[\Big(\int \sum_{j=1}^{q} a_T(j)\epsilon_{aq,t}\epsilon_{bq,t}\psi_{q,t-j}(v) \sum_{s=1}^{t-2q-1} \epsilon_{aq,s}\epsilon_{bq,s}\psi_{q,s-j}(v)^* dW(v)\Big)^2\Big] \\
&= \sum_{j=1}^{q}\sum_{l=1}^{q} a_T(j)a_T(l) \sum_{a,b=1}^{d} \iint \sum_{t=2q+2}^{T} \sum_{s=1}^{t-2q-1} \mathrm{E}\big[\epsilon_{aq,t}\epsilon_{bq,t}\psi_{q,t-j}(v)\psi_{q,t-l}(u)\big] \\
&\times \big[\epsilon_{aq,s}\epsilon_{aq,s}\psi_{q,s-l}(v)^*\psi_{q,s-j}(u)^*\big] dW(v)dW(u) \\
&= \frac{1}{2}\sum_{j=1}^{q}\sum_{l=1}^{q} k^2(j/h)k^2(l/p) \sum_{a,b=1}^{d} \iint \Big|\mathrm{E}\big[\epsilon_{aq,0}\epsilon_{bq,0}\psi_{q,-j}(v)\psi_{q,-l}(u)\big]\Big|^2 dW(v)dW(u)[1+o(1)].
\end{aligned}$$

Hence,

$$\mathrm{Var}(2\mathrm{Re}\tilde{V}_q) = 2\sum_{j=1}^{q}\sum_{l=1}^{q} k^2(j/h)k^2(l/p) \sum_{a,b=1}^{d} \iint \Big|\mathrm{E}\big[\epsilon_{a0}\epsilon_{b0}\psi_{-j}(v)\psi_{-l}(u)^*\big]\Big| dW(v)dW(u)[1+o(1)] \tag{C.30}$$

using $\mathrm{E}\big[\epsilon_{aq,0}\epsilon_{bq,0}\psi_{q,-j}(v)\psi_{q,-l}(u)^*\big] \to \mathrm{E}\big[\epsilon_{a0}\epsilon_{b0}\psi_{-j}(v)\psi_{-l}(u)^*\big]$ as $q \to \infty$, given Assumption A2. Put $C(0,j,l) \equiv \Big[(\epsilon_{a0}\epsilon_{b0} - \sigma_{ab})\psi_{-j}(v)\psi_{-l}(u)\Big]$, where $\sigma_{ab} = \mathrm{E}[\epsilon_{at}\epsilon_{bt}]$. Then

$$E\Big[\epsilon_{a0}\epsilon_{b0}\psi_{-j}(v)\psi_{-l}(u)\Big] = C(0,j,l) + \sigma_{ab}\sigma_{l-j}(v,u)$$

$$\begin{aligned}
\Big|\mathrm{E}[\epsilon_{a0}\epsilon_{b0}\psi_{-j}(v)\psi_{-l}(u)]\Big|^2 &= \Big|C(0,j,l)\Big|^2 + \sigma_{ab}^2\Big|\sigma_{l-j}(v,u)\Big|^2 \\
&+ 2\sigma_{ab}\mathrm{Re}\Big[C(0,j,l)\sigma_{l-j}^*(v,u)\Big].
\end{aligned}$$

Given that $\sum_{j=-\infty}^{\infty}\sum_{-\infty}^{\infty}|C(0,j,l)| \le C$, and $|k(\cdot)| \le 1$, we have

$$\begin{aligned}
\mathrm{Var}(2\mathrm{Re}\tilde{V}_q) &= 4\sum_{a,b=1}^{d}\sigma_{ab}^2\sum_{j=1}^{q}\sum_{l=1}^{q}k^2(j/h)k^2(l/p)\iint\left|\sigma_{l-j}(v,u)\right|^2 dW(v)dW(u)[1+o(1)] \\
&= 2\sum_{a,b=1}^{d}\sigma_{ab}^2 h\sum_{m=1-q}^{q-1}\Big[h^{-1}\sum_{j=m+1}^{q}k^2(j/h)k^2[(j-m)/h]\Big] \\
&\times\iint|\sigma_m(v,u)|^2 dW(v)dW(u)[1+o(1)] \\
&= 2\sum_{a,b=1}^{d}\sigma_{ab}^2 p\int_0^{\infty}k^4(z)\sum_{m=-\infty}^{\infty}\iint|\sigma_m(v,u)|^2 dW(v)dW(u)[1+o(1)] \\
&= 2\pi\sum_{a,b=1}^{d}\sigma_{ab}^2 p\int_0^{\infty}k^4(z)dz\iiint_{-\pi}^{\pi}|f(w,v,u)|^2 dw dW(v)dW(u)[1+o(1)]
\end{aligned}$$

using the fact that for any given m, $h^{-1}\sum_{j=m+1}^{q}k^2(j/h)k^2(\frac{j-m}{p}) \to \int_0^{\infty}k^4(z)dz$ as $p\to\infty$, and $q/p\to 0$.

We now verify condition (C.28). Noting that $\mathrm{E}|H_{j,t-2q-1}(v)|^4 \le Ct^2$ for $1\le j\le q$ given the MDS property of $\{\epsilon_{q,t},\mathcal{F}_{t-1}\}$ and Rosenthal's inequality (Hall and Heyde, 1980), we have

$$\begin{aligned}
\mathrm{E}|V_q(t)|^4 &\le \Big[\sum_{a=1}^{d}\sum_{j=1}^{q}a_T(j)\int\Big(\mathrm{E}|\epsilon_{aq,t}\psi_{q,t-2q-1}(v)H_{j,t-2q-1}(v)|^4\Big)^{\frac{1}{4}}dW(v)\Big]^4 \\
&\le Ct^2\Big[\sum_{j=1}^{q}a_T(j)\Big]^4 = O(h^4t^2/T^4).
\end{aligned}$$

It follows that $\sum_{t=2q+2}^{T}\mathrm{E}|V_q(t)|^4 = O(h^4/T) = o(h^2)$ given $h^2/T\to 0$. Thus (C.28) holds.

Next, we verify condition (C.29). Put $\sigma_{abq,t} \equiv \mathrm{E}[\epsilon_{aq,t}\epsilon_{bq,t}|\mathcal{F}_{t-1}]$. Then

$$
\begin{aligned}
\mathrm{E}[V_q^2(t)|\mathcal{F}_{t-1}] &= \sum_{a,b=1}^{d} \sigma_{abq,t}\Big[\sum_{j=1}^{q} a_T(j)\int \psi_{q,t-j}(v)H_{j,t-2q-1}(v)\Big]^2 \\
&= \sum_{a,b=1}^{d}\sum_{j,l=1}^{q} a_T(j)a_T(l)\iint \sigma_{abq,t}\psi_{q,t-j}(v)\psi_{q,t-l}(u) \\
&\times H_{j,t-2q-1}(v)H_{l,t-2q-1}(u)dW(v)dW(u) \\
&= \sum_{a,b=1}^{d}\sum_{j,l=1}^{q} a_T(j)a_T(l)\iint \mathrm{E}\big[\sigma_{abq,t}\psi_{q,t-j}(v)\psi_{q,t-l}(u)\big] \\
&\times H_{j,t-2q-1}(v)H_{l,t-2q-1}(u)dW(v)dW(u) \\
&+ \sum_{a,b=1}^{d}\sum_{j,l=1}^{q} a_T(j)a_T(l)\iint \tilde{Z}_{abq,t}^{j,l}(v,u)H_{j,t-2q-1}(v)H_{l,t-2q-1}(u)dW(v)dW(u) \\
&\equiv S_{1q}(t) + V_{1q}(t),
\end{aligned}
\tag{C.31}
$$

where $\tilde{Z}_{abq,t}^{j,l}(v,u) \equiv \sigma_{abq,t}\psi_{q,t-j}(v)\psi_{q,t-l}(u) - \mathrm{E}[\sigma_{abq,t}\psi_{q,t-j}(v)\psi_{q,t-l}(u)]$. Further decompose

$$
\begin{aligned}
S_{1,q} &= \sum_{a,b=1}^{d}\sum_{j,l=1}^{q} a_T(j)a_T(l)\int \mathrm{E}[\sigma_{abq,t}\psi_{q,t-j}(v)\psi_{q,t-l}(u)]\mathrm{E}[H_{l,t-2q-1}(v)H_{l,t-2q-1}(u)]dW(v)dW(u) \\
&+ \sum_{a,b=1}^{d}\sum_{j,l=1}^{q}\int \mathrm{E}[\sigma_{abq,t}\psi_{q,t-j}(v)\psi_{q,t-l}(u)] \\
&\times \Big\{H_{j,t-2q-1}(v)H_{l,t-2q-1}(u) - \mathrm{E}[H_{j,t-2q-1}(v)H_{l,2-2q-1}(u)]\Big\}dW(v)dW(u) \\
&\equiv \mathrm{E}[V_q^2(t)] + S_{2q}(t),
\end{aligned}
\tag{C.32}
$$

say, where

$$
\mathrm{E}[V_q^2(t)] = \sum_{a,b=1}^{d}\sum_{j,l=1}^{q}(t-q-1)a_T(j)a_T(l)\int \big|\mathrm{E}[\sigma_{abq,t}\psi_{q,t-j}(v)\psi_{q,t-l}(u)]\big|dW(v)dW(u)
$$

Then write

$$
\begin{aligned}
S_{2q}(t) = & \sum_{a,b=1}^{d}\sum_{j,l=1}^{q} a_T(j)a_T(l)\int \mathrm{E}[\sigma_{abq,t}\psi_{q,t-j}(v)\psi_{q,t-l}(u)] \\
& \times \sum_{s=1}^{t-2q-1} Z_{abq,s}^{j,l}(v,u)dW(v)dW(u) \\
& + \sum_{a,b=1}^{d}\sum_{j,l=1}^{q} a_T(j)a_T(l)\int \mathrm{E}[\sigma_{abq,t}\psi_{q,t-j}(v)\psi_{q,t-l}(u)] \\
& \times \sum_{s=2}^{t-2q-1}\sum_{\tau=1}^{s-1} \epsilon_{aq,s}\epsilon_{bq,\tau}\psi_{q,s-j}(v)\psi_{q,\tau-l}(u)dW(v)dW(u) \\
& \equiv V_{2q}(t) + S_{3q}(t), \text{say},
\end{aligned}
\tag{C.33}
$$

where

$$
\begin{aligned}
S_{3q}(t) = & \sum_{a,b=1}^{d}\sum_{j,l=1}^{q} a_T(j)a_T(l)\int \mathrm{E}[\sigma_{abq,t}\psi_{q,t-j}(v)\psi_{q,t-l}(u)] \\
& \times \sum_{0<s-\tau\le 2q}\ \sum_{0<s-\tau\le 2q} \epsilon_{aq,s}\epsilon_{bq,\tau}\psi_{q,s-j}(v)\psi_{q,\tau-l}(u)dW(v)dW(u) \\
& + \sum_{a,b=1}^{d}\sum_{j,l=1}^{q} a_T(j)a_T(l)\int \mathrm{E}[\epsilon_{aq,t}\epsilon_{bq,t}\psi_{q,t-j}(v)\psi_{q,t-l}(u)] \\
& \times \sum_{s-\tau>2q}\ \sum_{s-\tau>2q} \epsilon_{aq,s}\epsilon_{bq,\tau}\psi_{q,s-j}(v)\psi_{q,\tau-l}(u) \\
& \equiv V_{3q}(t) + V_{4q}(t), \text{say}.
\end{aligned}
\tag{C.34}
$$

Then, (C.29) follows from Lemmas C12-C15 below, which imply

$$
\mathrm{E}\Big|\sum_{t=2q+2}^{T} \mathrm{E}[V_q^2(t)|\mathcal{F}_{t-1}] - \mathrm{E}[V_q^2(t)]\Big|^2 = o(h^2)
$$

given $q = h^{1+\frac{1}{4b-2}}(\ln^2 T)^{\frac{1}{2b-1}}$ and $h = cT^{\lambda}$ for $0 < \lambda < (3 + \frac{1}{4b-2})^{-1}$. Thus, condition (C.29) holds, and so $M_q \xrightarrow{d} N(0,1)$ by Brown's theorem.

Lemma C12: Let $V_{1q}(t)$ be defined as in (C.31). Then $\mathrm{E}|\sum_{2q+2}^{T} V_{1q}(t)|^2 = O(qh^4/T)$.

Lemma C13: Let $V_{2q}(t)$ be defined as in (C.33). Then $\mathrm{E}|\sum_{2q+2}^{T} V_{2q}(t)|^2 = O(qh^4/T)$.

Lemma C14: Let $V_{3q}(t)$ be defined as in (C.34). Then $\mathrm{E}|\sum_{2q+2}^{T} V_{3q}(t)|^2 = O(qh^4/T)$.

Lemma C15: Let $V_{4q}(t)$ be defined as in (C.34). Then $\mathrm{E}|\sum_{2q+2}^{T} V_{4q}(t)|^2 = O(qh^4/T)$.

Proof of Lemma C12: Recall the definition of $\tilde{Z}^{j,l}_{abq,t}(v,u)$ as in (C.31). Noting that $\tilde{Z}^{j,l}_{abq,t}$ is independent of $\{H_{j,t-2q-1}(v)H_{l,t-2q-1}(u)\}$ and that $\tilde{Z}^{j,l}_{abq,t}(v,u)$ is independent of $\tilde{Z}^{j,l}_{abq,\tau}(v,u)$ for $t-\tau > 2q$ and $1 \le j,l \le q$, we can obtain

$$
\begin{aligned}
\mathrm{E}\Big|\sum_{a,b=1}^{d}\sum_{t=2q+2}^{T}\tilde{Z}^{j,l}_{abq,t}(v,u)H_{j,t-2q-1}(v)H_{l,t-2q-1}(a)\Big|^2 \le{}& \sum_{a,b=1}^{d}\sum_{|t-\tau|\le 2q}\sum_{|t-\tau|\le 2q}\mathrm{E}|\tilde{Z}^{j,l}_{abq,t}(v,u)\tilde{Z}^{j,l}_{abq,\tau}(v,u)| \\
&\times\Big(\mathrm{E}|H_{j,t-2q-1}(v)|^4\Big)^{\frac{1}{4}}\Big(\mathrm{E}|H_{l,t-2q-1}(u)|^4\Big)^{\frac{1}{4}} \\
&\times\Big(\mathrm{E}|H_{j,\tau-2q-1}(v)|^4\Big)^{\frac{1}{4}}\Big(\mathrm{E}|H_{l,\tau-2q-1}(u)|^4\Big)^{\frac{1}{4}} \\
={}& O(T^3q),
\end{aligned}
$$

using the fact that $\mathrm{E}|H_{j,\tau-2q-1}(v)|^4 \le Ct^2$ for $1 \le j \le q$. It follows by Minkowski's inequality and (C.2) that

$$
\begin{aligned}
\mathrm{E}|\sum_{t=2q+2}^{T}V_{1q}(t)|^2 \le{}& \Big[\sum_{a,b=1}^{d}\sum_{j,l=1}^{T}a_T(j)a_T(l) \\
&\times\Big(\mathrm{E}|\sum_{2q+2}^{T}\iint\tilde{Z}^{j,l}_{abq,t}(v,u)H_{j,t-2q-1}(v)H_{l,t-2q-1}(u)dW(v)dW(u)|^2\Big)^{\frac{1}{2}}\Big]^2 \\
={}& O(qh^4/T).
\end{aligned}
$$

□

Proof of Lemma C13: Recalling the definition of $Z^{j,l}_{abq,s}(v,u)$ in (C.33) and noting that $\{Z^{j,l}_{abq,s}(v,u)\}_{j,l=1}^{q}$ is independent of $\{Z^{j,l}_{abq,\tau}(v,u)\}_{j,l=1}^{q}$ for $|s-\tau| > 2q$ where q is sufficiently large, we have

$$
\Big|\sum_{s=1}^{t-q-1}Z^{j,l}_{abq,s}(v,u)\Big|^2 = \sum\sum_{|s-\tau|\le 2q}\mathrm{E}\Big[Z^{j,l}_{abq,s}(v,u)Z^{j,l}_{abq,\tau}(v,u)\Big] \le 2Ctq.
$$

It follows that

$$
\begin{aligned}
\mathrm{E}\Big|\sum_{a,b=1}^{d}\sum_{t=2q+2}^{T}V_{2q}(t)\Big|^2 \le{}& \Big\{\sum_{a,b=1}^{d}\sum_{t=2q+2}^{T}\Big[\mathrm{E}|V_{2q}(t)|^2\Big]^{\frac{1}{2}}\Big\}^2 \\
\le{}& \Big\{\sum_{a,b=1}^{d}\sum_{t=2q+2}^{2}\sum_{j,l=1}^{q}a_T(j)a_T(l)\int\Big|\mathrm{E}\big[\sigma_{abq,t}\psi_{q,t-j}(v)\psi_{q,t-l}(u)\big]\Big| \\
&\times\Big(\mathrm{E}|\sum_{s=1}^{t-2q-1}Z^{j,l}_{abq,s}(v,u)|^2\Big)^{\frac{1}{2}}dW(v)dW(u)\Big\}^2 \\
={}& O(qh^4/T).
\end{aligned}
$$

□

Proof of Lemma C14: From Minkowski's inequality and

$$\begin{aligned}
\mathrm{E}|V_{3q}(t)|^2 \leq & \Big[\sum_{a,b=1}^{d} \sum_{j,l=1}^{q} a_T(j) a_T(l) \int |\mathrm{E}[\epsilon_{aq,t}\epsilon_{bq,t}\psi_{q,t-j}(v)\psi_{q,t-l}(u)]| \\
& \times \Big[\sum_{s=1}^{t-2q-1} \mathrm{E}\Big|\epsilon_{aq,s}\psi_{q,s-j}(v) \sum_{s-\tau \leq 2q} \epsilon_{bq,\tau}\psi_{q,\tau-l}(u)\Big|^2\Big]^{\frac{1}{2}} dW(v)dW(u)\Big]^2 \\
& \leq 2Ctq\Big[\sum_{j=1}^{q} a_T(j)\Big]^4 = O(tqh^4/T^4),
\end{aligned}$$

we have $\mathrm{E}|\sum_{t=2q+2}^{T} V_{3q}(t)|^2 = O(qh^2/T)$.□

Proof of Lemma C15: The proof follows from $\sum_{j,l=1}^{\infty} \Big|\mathrm{E}[\big(\epsilon_{aq,0}\epsilon_{bq,0}-\sigma_{abq,t}\big)\psi_{q,t-j}(v)\psi_{q,t-l}(u)]\Big| \leq C$, Minkowski's inequality and

$$\begin{aligned}
\mathrm{E}|V_{4q}(t)|^2 =& \mathrm{E}\Big| \sum_{a,b=1}^{d} \sum_{j,l=1}^{q} a_T(j)a_T(l) \int \mathrm{E}\Big[\epsilon_{aq,0}\epsilon_{bq,0}\psi_{q,-j}(v)\psi_{q,-l}(u)\Big] \sum_{s=2q+2}^{t-2q-1} \epsilon_{bq,s}\psi_{q,s-j}(v) \\
& \times \sum_{\tau=1}^{s-2q-1} \epsilon_{aq,\tau}\psi_{q,\tau-l}(u)dW(v)dW(u)\Big|^2 \\
=& \sum_{j_1,l_1=1}^{q} \sum_{j_2,l_2=1}^{q} a_T(j_1)a_T(j_2)a_T(l_1)a_T(l_2) \\
& \times \iint\iint \mathrm{E}[\epsilon_{aq,0}\epsilon_{bq,0}\psi_{q,-j_1}(v_1)\psi_{q,-l_1}(u_1)]\mathrm{E}[\epsilon_{aq,0}\epsilon_{bq,0}\psi^*_{q,-j_2}(v_2)\psi^*_{q,-l_2}(u_2)] \\
& \times \sum_{s=2q+2}^{t-2q-1} \mathrm{E}[\epsilon_{aq,s}\epsilon_{bq,\tau}\psi_{q,s-j_1}(v_1)\psi_{q,s-j_2}(v_2)] \\
& \times \sum_{\tau=1}^{s-2q-1} \mathrm{E}[\epsilon_{aq,\tau}\epsilon_{bq,\tau}\psi^*_{q,\tau-l_1}(u_1)\psi^*_{q,\tau-l_2}(u_2)] \\
& \times dW(v_1)dW(u_1)dW(v_2)dW(u_2) = O(t^2h/T^4).
\end{aligned}$$

□

Proof of Theorem 4.1(b): The proof of Theorem 4.1(b) consists of the proofs of Theorems C4 and C5 below.

Theorem C4: Under the conditions of Theorem 4.1(b), $(\frac{h^{\frac{1}{2}}}{T})[\hat{M} - M] \xrightarrow{p} 0$.

Theorem C5: Under the conditions of Theorem 4.1(b),

$$(\frac{h^{\frac{1}{2}}}{T})M \xrightarrow{p} (2D)^{-\frac{1}{2}}\pi \iint_{-\pi}^{\pi} \left\| f^{(0,1,0)}(w,0,v) - f_0^{(0,1,0)}(w,0,v)\right\|^2 dwdW(v)$$

Proof of Theorem C4: It suffices to show that

$$T^{-1}\int \sum_{j=1}^{T-1} k^2(j/h)T_j\Big[\|\hat{\sigma}_j^{(1,0)}(0,v)\|^2 - \|\tilde{\sigma}_j^{(1,0)}(0,v)\|^2\Big]dW(v) \xrightarrow{p} 0 \tag{C.35}$$

$h^{-1}[\hat{C}-\tilde{C}] = O_P(1)$, $h^{-1}[\hat{D}-\tilde{D}] \xrightarrow{p} 0$, where $\tilde{C}$ and $\tilde{D}$ are defined in the same way as $\hat{C}$ and $\hat{D}$, with $\{\epsilon_t\}_{t=1}^T$ replacing $\{\hat{\epsilon}_t\}_{t=1}^T$. I only prove (C.35). From (C.2), the Cauchy-Schwarz inequality, and the fact that

$$T^{-1}\int \sum_{j=1}^{T-1} k^2(j/h)T_j\|\tilde{\sigma}_j^{(1,0)}(0,v)\|^2 dW(v) = O_P(1),$$

it suffices to show that $T^{-1}\hat{A}_1 \xrightarrow{p} 0$, where $\hat{A}_1$ is defined as in (C.7). Given (C.8), we shall show that $T^{-1}\int\sum_{j=1}^{T-1} k^2(j/h)T_j\|\hat{B}_{aj}(v)\|^2 dW(v) \xrightarrow{p} 0$, $a = 1,\cdots,6$. I first consider $a=1$. By the Cauchy-Schwarz inequality and $|\hat{\delta}_t(v)| \le 2$, we have

$$\begin{aligned} T^{-1}\int \sum_{j=1}^{T-1} k^2(j/h)T_j\|\hat{B}_{1j}(v)\|^2 dW(v) \le &\Big[\sum_{t=1}^{T}\|\hat{\epsilon}_t - \epsilon_t\|^2\Big] \\ &\times \sum_{j=1}^{T-1} a_T(j)\Big[\int dW(v)\Big]^2 = O_P(h/T) \end{aligned} \tag{C.36}$$

using (C.1), (C.2) and Assumption A6.

The proof for $a=2$ is similar, noting that

$$\|T_j^{-1}\sum_{t=j+1}^{T}(\hat{\epsilon}_t - \epsilon_t)\|^2 \le T_j^{-1}\sum_{t=j+1}^{T}\|\hat{\epsilon}_t - \epsilon_t\|^2.$$

For $a=3$, using Cauchy-Schwarz inequality, we have

$$\begin{aligned} T^{-1}\int \sum_{j=1}^{T-1} k^2(j/h)T_j\|\hat{B}_{3j}(v)\|^2 dW(v) \le &\Big(T^{-1}\sum_{t=1}^{T}\|\epsilon_t\|^2\Big)\Big[T^{-1}\sum_{t=1}^{T}\|\hat{\epsilon}_t - \epsilon_t\|^2\Big] \\ &\times \sum_{j=1}^{T-1} k^2(j/h)\int \|v\|^2 dW(v) \\ &= O_P(h/T) \end{aligned} \tag{C.37}$$

The proof for $a = 4, 5, 6$ is similar to that for $a = 3$, noting that $|T_j^{-1} \sum_{t=j+1}^{T} \hat{\delta}_t(v)|^2 \leq T_j^{-1} \sum_{t=j+1}^{T} |\hat{\delta}_t(v)|^2$. This completes the proof for Theorem C4. □

Proof of Theorem C5: The proof is similar to Hong (1999, Proof of Theorem 5), for the case $(m, l) = (1, 0)$ and $W_1(\cdot) = \delta(\cdot)$, the Dirac delta function. □

References

Brown, B. M., 1971. Martingale central limit theorems. *The Annals of Mathematical Statistics*, 42(1):59–66.

Hall, P. and Heyde, C. C., 1980. *Martingale limit theory and its application.* Academic press New York.

Hong, Y., 1999. Hypothesis testing in time series via the empirical characteristic function: A generalized spectral density approach. *Journal of the American Statistical Association*, 94 (448):1201–1220.

Hong, Y. and Lee, Y.-J., 2005. Generalized spectral tests for conditional mean models in time series with conditional heteroscedasticity of unknown form. *The Review of Economic Studies*, 72(2):499–541.

Ling, S. and McAleer, M., 2003. Asymptotic theory for a vector arma-garch model. *Econometric Theory*, 19(02):280–310.

News, Noise, and Tests of Present Value Models

Mehdi Hamidi Sahneh *

This paper develops a present value model of stock prices in which agents receive a noisy signal about the future economic fundamentals. The noisy signal serves two purposes. *First*, the noise in the signal generates temporary price fluctuations which are unrelated to economic fundamentals. *Second*, the noisy signal drives a wedge between the information set of agents and econometricians, which poses substantial challenges for econometric testing of market efficiency. By employing conventional econometric analyses, an outside econometrician unknowingly conditions on a smaller information set than the agents. As a result, he might find that stock prices are excessively volatile and wrongly reject the cross-equation restrictions implied by the underlying economic model. I show that rationality restricts the contribution of noise to price fluctuations. My estimates show that the US stock market was indeed reacting rationally to news, and the variation in price-dividend ratio is mostly driven by noise and movements in discount rates, and almost nothing from movements in expected dividend growth.

Keywords: Efficient Market Hypothesis; Cross-Equation Restrictions; Excess Volatility; Non-Causal VAR Representation.

JEL classification: C5, G12, D82

*The author is deeply indebted to Carlos Velasco for guidance and encouragement. I also benefited most from the comments of Fabio Canova, Luca Gambetti, Jesus Gonzalo, Matthias Kredler, Jose Penalva, and Hernan Seoanes, as well as to seminar participants at UC3m, UAB Barcelona, and Pompeu Fabra University.

1 Introduction

I develop a present value model of stock prices in which agents receive a noisy signal about the future economic fundamentals. The model's solution takes the form of a forward-looking (or non-causal) representation that poses substantial challenges for econometric estimation and inference. By employing conventional econometric analyses, such as vector autoregressions (VARs), an outside econometrician unknowingly conditions on a smaller information set than the agents and misspecifies the true dynamics of the equilibrium. In this framework, conventional empirical tests can find patterns in the data that are different from those perceived by rational investors.[1] In particular, I show that excess volatility and violation of cross-equation restrictions could be reconciled with the data, once we allow for non-causality. Finally, I show that market efficacy imposes an upper bound on the contribution of noise to price volatility. My find that, (1) we can not reject the null hypothesis of market efficiency; and (2) the US stock market fluctuations have been dominated by noise and movements in discount rates, and almost nothing from movements in dividend growth.

Although the insight of this article applies to many areas of economics that incorporate *Rational Expectation Hypothesis*, for several reasons I focus on the econometrics testing of stock market efficiency. First, given the crucial role of the financial sector in the modern economy, as is evident from the world-wide financial crisis of 2008-2009, the question of market efficiency is of great importance to both policy makers and academics.[2] Second, stock market provides almost ideal conditions for testing rationality and efficiency. There are few other areas of economics with a large central market, no entry cost and the possibility of short selling. Third, there is a huge financial incentive to analyze and react rationally to new information. Finally, detailed and precise information on historical

[1] "market participants", "agents", and "investors" are all the same here.

[2] Some have even argued that market efficiency and rationality was partly responsible for the financial crisis. For an accessible survey of the EMH see Malkiel (2011).

prices, dividends, earning, etc, is available.

Fama (1970) defines market efficiency as: "a market in which prices always 'fully reflect' (his emphasis) available information." Markets have a variety of sources for information. Specifically, we can not rule out that agents have some information about the future economic fundamentals. These sources include public release of information such as macroeconomic forecasts, economic surveys, policy makers' statements, and news about technological innovations (such as computer, Internet, etc).

To formalize this point, I introduce a noisy signal into a standard PV model with constant discount rate. The signal, however, is contaminated with noise, as news in the real world is not perfect. The noisy signal serves two purposes. *First*, the presence of noise in the signal gives rise to temporary fluctuations in stock prices which are unrelated to economic fundamentals. I show that my model can produce interesting asset market dynamics. In the data, some booms in asset prices are followed by busts and high prices are followed by low returns and vice versa. My model is consistent with these observations.

Second, the presence of the noisy signal about the future economic fundamentals generates a non-causal equilibrium solution. By employing a conventional VAR analysis, econometricians unknowingly condition on a smaller information set than the agents. Not surprisingly, this has some consequences for econometrics testing of market efficiency. In this framework, I show that an outside econometrician who only observes realized prices and dividends can find patterns in the data that are different from those perceived by rational agents. Specifically, an econometrician who employs Campbell-Shiller testing procedure, will find that stock prices display excess volatility and wrongly reject the null hypothesis of market efficiency. Although non-causal processes are well-known in the statistical and time series literature, no economic model has been presented that gives rise to such a representation. My framework is the first to do so.

Finally, I show that rationality restricts the contribution of noise to price volatility. Specifically, market efficiency implies that noise should not explain more than half of price fluctuation. The fact that noise variance has non-monotonic effects on the price's dynamics drives this result. This is because noise variance affects not only the volatility of the shocks, but also the inference problem of the investors. In particular, when this variance is either too small or too large, noise generates small price fluctuations. In the first case, signals are very precise. In the second case, signals are very imprecise and investors disregard them in their inference. The variance of noise component is largest for intermediate levels of noise variance.

In order to test this restriction, I propose a new procedure to decompose the US stock prices into a fundamental component and a noise component. Because noise is unrelated to future economic fundamentals, we can define noise as the component of price which is orthogonal to the future economic fundamentals. I argue that my decomposition is plausible because it closely matches the anecdotal evidence in the asset pricing literature. My approach does not require any assumption on the structure of the model or the information set of market participants. My estimate shows that while we can not reject the null hypothesis of market efficiency, I find that the U.S. stock market fluctuations have been dominated by news about future returns, not by news about future dividends. Moreover, I find that the US stock market was undervalued during the 1970s and overvalued during the 1990s, known as the *dot-com* bubble, but there is no evidence that the market was overvalued before the Wall Street Crash of 1929.

Relation to the Literature. This paper relates to the news in the business cycle literature. Beaudry and Portier (2006) provide empirical evidence supporting the view that news about the future is an important source of aggregate macroeconomic fluctuations. Leeper et al. (2013) show that foresight about future economic fundamentals can create equilibria with non-fundamental moving average representations. It is well-known that non-fundamentalness pose challenges to standard econometrics efforts to recover structural shocks and impulse response functions. Forni et al. (2016) propose a novel dynamic identification strategy to recover the structural shocks and impulse response functions from a non-fundamental VAR model.

Non-causality and non-fundamentalness are two sides of the same coin: non-fundamentalness arises when the moving average polynomial has a root inside the unit circle; instead, non-causality arises when the autoregressive polynomial has a root inside the unit circle. There is a rich literature in statistical time series suggesting that non-causal processes can display interesting dynamics often observed in economic and financial time series. For instance, Gouriéroux and Zakoïan (2016) show that non-causal autoregressive processes allow for local explosion or dynamics that may look like GARCH to an outside econometrician. However, to the best of my knowledge, no economic model has been presented that gives rise to such a representation. Therefore, a key contribution of this paper is to propose an economic model that gives rise to non-causal equilibrium representations.

This paper also relates to Kasa et al. (2014). These authors argue that if the asset markets feature persistent heterogeneous beliefs, an econometrician who incorrectly assumes agents have homogeneous beliefs might incorrectly reject the null hypothesis of market efficiency. Although heterogeneous beliefs models are very interesting, they have a reputation for being difficult to handle. To the contrary, non-causal representations can be tested and estimated.[3]

The remainder of the paper is organized as follows. Section II presents a simple analytical example around which the discussion is organized. Section III discusses the implication of non-causality for the tests of the present value model. In section IV, I propose a test to detect non-causality. Section V presents the results of the historical decomposition of US stock prices. The last Section provides concluding comments.

[3]Lanne and Saikkonen (2013) and Davis and Song (2012) propose estimation procedures for non-causal VAR models.

2 Analytical Example

Let me begin by considering a general one-period Euler equation:

$$P_t = \mathbb{E}\big[M_t(P_{t+1} + D_t)|I_t\big] \tag{2.1}$$

where P_t, is the asset price at the beginning of period t, D_t is the dividend paid at the end of period t, and M_t is the discount factor. Moreover, $\mathbb{E}$ denotes expectation conditional on all available information to the market participants at the beginning of period t, denoted by I_t.

To complete the description of the model, we must specify a process for the discount factor, dividends, and how agents make optimal forecasts based on their information set. Some of these assumptions are not realistic, but they provide a good starting point for my discussion. Nevertheless, the setting in this section could be generalized to allow for more realistic assumptions, such as time-varying discount factors, at the cost of a more complicated notation and probably analytical solution.

Discount Rates. For ease of exposition, I consider a linear utility function, which implies that $M_t = \beta$, for all t. A long tradition in asset pricing has opted for consumption-based discount factors, such as constant relative risk aversion (CRRA) utility functions. Linear utility function is equivalent to a CRRA utility function with the parameter of risk aversion equal to *zero*.

Dividends. I consider a typical unit root process

$$D_t = D_{t-1} + \epsilon_t, \tag{2.2}$$

where the dividend change, ϵ_t, follows an *independent and identically distributed* (*iid*) process with mean *zero* and variance σ_ϵ^2, that is $\epsilon_t \sim iid(0, \sigma_\epsilon^2)$. Although extensive empirical evidence support this assumption, my setting can accommodate both stationary and nonstationary ARMA processes.

Agents' Information Set. I assume that agents observe the complete history of prices and dividends. Moreover, at the beginning of each period and right before prices are determined, agents observe some noisy signals about the next period dividend change. Following the notation that $s_{t|t+1}$ denotes the news at time t about time $t+1$ dividend, we have

$$s_{t|t+1} = \epsilon_{t+1} + \nu_t, \tag{2.3}$$

where $\nu_t \sim iid(0, \sigma_\nu^2)$ denotes the noise which is orthogonal to ϵ_t at all leads and lags. Thus, the agents' information set at the beginning of period t (say I_t), encompass the current and past signals and dividends, as well as the history of prices.

A shortcoming of such specification is that it requires a priori restriction that agents do not observe signals beyond one period. In Appendix (A), I generalize the signal specification, allowing agents to observe a noisy signal about the entire future dividend changes.

Expectations. I have not yet specified how agents make optimal predictions. For convenience, I assume that agents' optimal forecast is linear.[4] Therefore, $\mathbb{E}[D_{t+j}|I_t]$, for $j = 1, 2, \cdots$, is simply the linear projection of D_{t+j} on I_t, which implies that

$$\begin{aligned} \mathbb{E}[D_{t+j}|I_t] &= D_t + \gamma s_{t|t+1}, \\ &= D_t + \gamma(\epsilon_{t+1} + \nu_t), \; j = 1, 2, \cdots \end{aligned} \tag{2.4}$$

where $\gamma = \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + \sigma_\nu^2}$.

[4]The linear optimal forecast is the standard assumption in present value models. See, for instance, Campbell and Shiller (1987, 1988a,b). See Donaldson and Kamstra (1996) for an exception. For a more detailed derivation of linear projection, the reader should consult Hamilton (1994), Ch. 4.

We can now find equilibrium stock prices. Solving (2.1) forward, together with the transversality condition we obtain a standard Present Value (PV) model with constant discount factor

$$P_t = \sum_{j=0}^{\infty} \beta^{j+1} \mathbb{E}(D_{t+j}|I_t), \tag{2.5}$$

where in the asset pricing literature the right hand side is known as *fundamental value*, and any deviation of price from fundamental values is known as *bubble*. For an accessible survey of the bubble literature see Brunnermeier (2009).

Substituting (2.4) into (2.5), the generalized form of equilibrium price is given by

$$\begin{aligned} P_t &= \kappa D_t + \kappa\beta\gamma s_{t|t+1} \\ &= \underbrace{\kappa D_t + \kappa\beta\gamma\epsilon_{t+1}}_{fundamental\ component} + \underbrace{\kappa\beta\gamma\nu_t}_{noise\ component}, \end{aligned} \tag{2.6}$$

where $\kappa = \frac{\beta}{1-\beta}$. The intuition behind (2.6) is simple. Rational agents observe the news which provides them with some noise information about the future economic fundamentals. Since the agents do not know exactly which part of the signal is noise, they are cautions and discount the news by parameter $\gamma < 1$. And they discount the news more if the news is more imprecise, i.e. $\mathrm{var}(\sigma_\nu^2)$ is large.

From (2.6) we see that the stock price can be decomposed into two components, a fundamental component ($\kappa D_t + \kappa\beta\gamma\epsilon_{t+1}$), associated with the future dividends, and a component associated with the noise ($\kappa\beta\gamma\nu_t$). Note that both fundamental component and noise component are part of the fundamental value.

2.1 Testable Restrictions

In this section, I discuss some interesting properties of the noise component and the restrictions that rationality imposes on the time series of the data. The size of the noise component is controlled by the variance of the noise. The fact that noise variance affects not only the volatility of the shocks, but also the inference problem of the investors drives this result.

Two interesting limit cases are $\sigma_\nu^2 \to 0$, i.e. the signal is precise, and $\sigma_\nu^2 \to \infty$, i.e. the signal is largely noise. When $\sigma_\nu^2 \to 0$, the signal is accurate and agents can see the future dividend changes. As a result, the variance of the noise component, $\frac{\sigma_\epsilon^4 \sigma_\nu^2}{(\sigma_\nu^2+\sigma_\epsilon^2)^2}$, vanishes when $\sigma_\nu^2 \to 0$ and there is no noise component. Interestingly, the noise component disappears even in the opposite case, when $\sigma_\nu^2 \to \infty$. To see this point, note that the variance of the noise component also vanishes when $\sigma_\nu^2 \to \infty$. The economic intuition is that, when ν_t is very large, the signal is not informative, so that agents ignore it.

The noise component is large when $\sigma_\nu^2 = \sigma_\epsilon^2$. This can be seen from the ratio of the variance of the noise component to the variance of prices, $\frac{\sigma_\nu^2}{\sigma_\nu^2+\sigma_\epsilon^2}$, which reaches its maximum 1/2 when the variance of the noise equals the variance of dividend change. Therefore, rationality implies that noise component can not explain more than 50% of price fluctuations.

On the other hand, irrational investors would not discount the noisy signal, i.e. $\gamma = 1$, which implies that the price deviates from the fundamental value, which I refer to as *noise bubble*, to emphasize the difference with the *noise component*. Therefore, one can detect irrationality by testing if noise explains more than 50% of price fluctuations.

2.2 The Econometrics of Non-Causality

In this section, I will argue that an econometrician who uses standard econometrics techniques such as VAR, *unknowingly* misspecifies the true dynamics of the model and conditions on a smaller information set than agents.

For the moment, suppose the signal is accurate (i.e. $\gamma = 1$), and $\kappa = 1$. Then we can write (2.6) as

$$
\begin{aligned}
S_t &= \epsilon_t + \beta \epsilon_{t+1} \\
&= (1 + \beta F)\epsilon_t, \qquad (2.7)
\end{aligned}
$$

where $S_t \equiv P_t - \kappa D_{t-1}$ is the spread between prices and a multiple of dividends, and $F = L^{-1}$ is the forward operator (*i.e.*, $FX_t = X_{t+1}$). The moving average terms that the signal produces pose substantial challenges for econometric inference. Representation (2.7) is invertible in current and future spread:

$$
\epsilon_t = \frac{S_t}{1 + \beta F} = S_t - \beta S_{t+1} + \beta S_{t+2} - \cdots .
$$

An econometrician who uses conventional econometrics analysis, *unknowingly* estimates the following backward-looking process

$$
\begin{aligned}
S_t &= \underbrace{(1 + \beta F)\left[\frac{1 + \beta L}{1 + \beta F}\right]}\ \underbrace{\left[\frac{1 + \beta F}{1 + \beta L}\right]\epsilon_t} \\
&= \qquad (1 + \beta L) \qquad\qquad \tilde{\epsilon}_t \\
&= \tilde{\epsilon}_t + \beta \tilde{\epsilon}_{t-1} \qquad (2.8)
\end{aligned}
$$

where the econometrician's innovations, $\tilde{\epsilon}_t = \frac{1+\beta F}{1+\beta L}\epsilon_t$, are the statistical shocks obtained from linear projections of S_t on its past

$$
\tilde{\epsilon}_t = \frac{S_t}{1 + \beta L} = S_t - \beta S_{t-1} + \beta^2 S_{t-2} - \cdots . \qquad (2.9)
$$

Representation (2.8) is derived by flipping the root of the MA polynomial from inside to outside the unit circle via the Blaschke factor, $\frac{1+\beta F}{1+\beta L}$. See Hansen et al. (1981) and Lippi and Reichlin (1994) for more details on Blaschke factors. This misspecification is due to the fact that (2.7) and (2.8) have the same covariance structure, and estimation procedures that use up to second-order moments of the data can not identify causal and non-causal systems. Not surprisingly, this misspecification has some consequences.

The econometrician's information set (H_t) corresponds to the linear space spanned by current and past S_t, where from the Wold representation, is equivalent to the linear space spanned by the current and past values of $\tilde{\epsilon}_t = \frac{1+\beta F}{1+\beta L}\epsilon_t$. Since the Blaschke factor is a two-sided filter, each element $\tilde{\epsilon}_t$ is a function of past, current, and future values of ϵ_t

$$
\begin{aligned}
\tilde{\epsilon}_t &= (1+\beta F)\sum_{j=0}^{\infty} -\beta^j \epsilon_{t-j} \\
&= \beta\epsilon_{t+1} + (1-\beta^2)\epsilon_t - \beta(1-\beta^2)\epsilon_{t-1} + \beta^2(1-\beta^2)\epsilon_{t-2} - \beta^3(1-\beta^2)\epsilon_{t-3} + \cdots .
\end{aligned}
$$

As a result, the closed linear space generated by current and past values of $\tilde{\epsilon}_t$ is no larger than the linear space generated by ϵ_t. Thus, an econometrician who uses standard econometrics techniques such as VAR, *unknowingly* conditions on a smaller information set than agents, i.e. $H_t \subseteq I_t$.

Augmenting the econometrician's information set by dividends does not solve the information misalignment. To see this point, suppose the econometrician estimates a VAR that includes the first difference of dividends ΔD_t and the spread. The advantage of conditioning on $Z_t \equiv (S_t, \Delta D_t)$ instead of first differences of prices and dividends is that Z_t summarize the joint history of prices and dividends, and at the same time does not loose information on the levels of these variables (Campbell and Shiller, 1987). The equilibrium solution for the dividends and the spread is given by

$$
\begin{bmatrix} 1-L & 0 \\ -\kappa L & 1 \end{bmatrix} \begin{bmatrix} D_t \\ P_t \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \kappa\beta\gamma & \kappa(1+\beta\gamma F) \end{bmatrix} \underbrace{\begin{bmatrix} \nu_t \\ \epsilon_t \end{bmatrix}}_{\eta_t} \tag{2.10}
$$

which is a forward-looking representation.

By employing standard econometrics procedures, the econometrician *unknowingly* estimates a backward-looking model

$$\begin{bmatrix} 1-L & 0 \\ -\kappa L & 1 \end{bmatrix} \begin{bmatrix} D_t \\ P_t \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1+\beta\gamma L \end{bmatrix} \underbrace{\begin{bmatrix} 0 & 1 \\ \frac{\kappa\beta\gamma}{(1+\beta\gamma L)} & \frac{\kappa(1+\beta\gamma F)}{(1+\beta\gamma L)} \end{bmatrix} \begin{bmatrix} \nu_t \\ \epsilon_t \end{bmatrix}}_{\tilde{\eta}_t} \tag{2.11}$$

where (2.11) now has a VAR representation, mapping $\tilde{\eta}_t$ into current and past values of the observables. The econometrician's information set is the linear space spanned by the current and past values of the observables, $Z_t \equiv (\Delta D_t, S_t)$, which corresponds to the linear space spanned by the current and past values of $\tilde{\eta}_t$. Since each element of $\tilde{\eta}_t$ is a linear function of past, current, and future values of η_t, the closed linear space generated by current and past values of $\tilde{\eta}_t$ is no larger than the space generated by the η_t, which spans the agents' information set.

3 Tests of Present Value Models

I now argue that an outside econometrician who only observes realized prices and dividends can find patterns in the data that are different from those perceived by agents. Errors arise because by employing standard econometrics techniques, an econometrician misspecifies the dynamics of the model and the information set of market participants will be larger than the econometricians.

3.1 Excess Volatility

3.1.1 Correspondence between P_t and P_t^*

In his seminal paper, Shiller (1981) contrasts the plots of prices, P_t, with its ex-post rational counterpart, P_t^*, defined as

$$P_t^* = \sum_{j=0}^{\infty} \beta^j D_{t+j}, \tag{3.1}$$

and argues that stock prices move too much to be justified by future dividends. Motivated by the joint hypothesis problem, many criticized the constant discount factor assumption. In response, Grossman and Shiller (1981) consider a CRRA utility function,

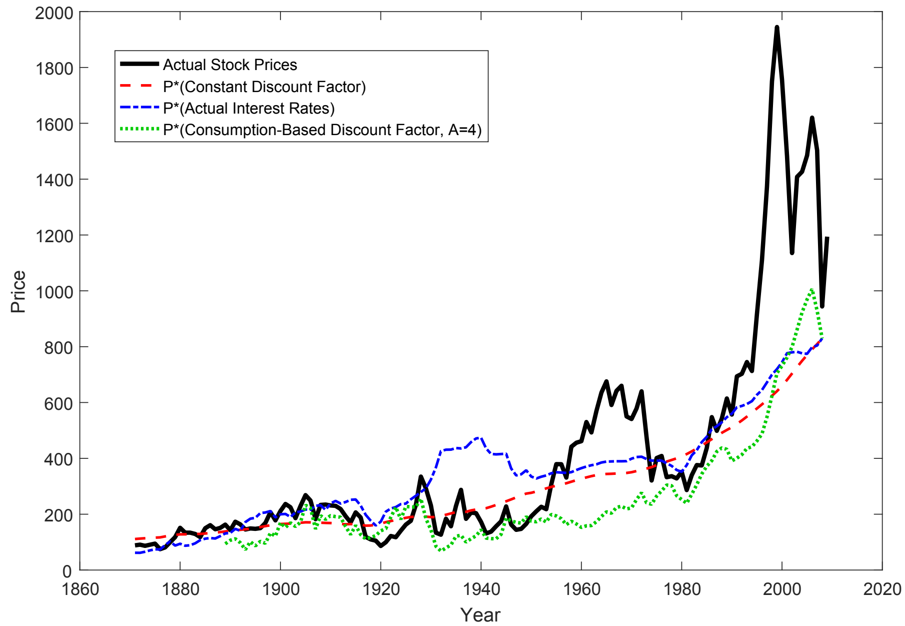
$$U(C) = \frac{1}{1-A} C^{1-A}, \qquad 0 < A < \infty,$$

where A is the coefficient of relative risk aversion, which is a measure of the concavity of the utility function, and C is the aggregate consumption. Thus, price equals the expected present value of dividends discounted by the marginal rates of substitution:

$$P_t = \mathbb{E}\Big[\sum_{j=1}^{\infty} \beta^j \frac{U'(C_{t+j})}{U'(C_t)} D_{t+j} | I_t\Big].$$

Figure (1) plots the real stock prices, along with three different perfect foresight measures that differ from each other only in the assumed discount rates. All data have been obtained from Shiller's website. These plots give the impression that stock prices deviate from the fundamental values, so much so that in his Noble lecture, Shiller (2014) concludes: "Once again, the figure reveals that there is little correspondence between any of these measures of ex-post rational price and actual stock price. People did not behave, in setting stock

Figure 1: Actual Real Stock Price with Three Alternative PDVs of Future Real Dividends



Notes: Real S&P Stock Index, along with present value of future dividends discounted by constant, interest rate and consumption based discount factors, 1871-2009.

prices."[5]

Figure (2) compares the sample path of simulated stock prices generated by model (2.10). To produce Figure (2), I generate $\epsilon_t \sim iid\ N(0, \sigma_\epsilon^2)$, and $\nu_t \sim iid\ N(0, \sigma_\nu^2)$. The parameter values $\mu = 0.1881$, $\sigma_\epsilon = 1.376$, and $d_0 = 4.7314$ are set to correspond to estimates for Standard and Poor's (deflated) annual dividend and price series from 1871 to 2012. I also set $\sigma_\nu^2 = \sigma_\epsilon^2$ and $\beta = 0.96$ to match the average annual real interest rate in the sample period.

Similar characteristics are apparent in the simulated data generated by my model. Therefore, the lack of correspondence between P_t and P_t^* can not be interpreted as evidence against the PV model or rationality. The intuition is simple: P_t does not correspond to P_t^* if investors have noisy information about the future dividends and discount factors,

[5]Using Monte Carlo simulations, Kleidon (1986) found similar patterns for nonstationary dividends and rationally determined stock prices, and concludes that plots of P_t and P_t^* should not be used as evidence of excess volatility or against constant discount factor.

whi

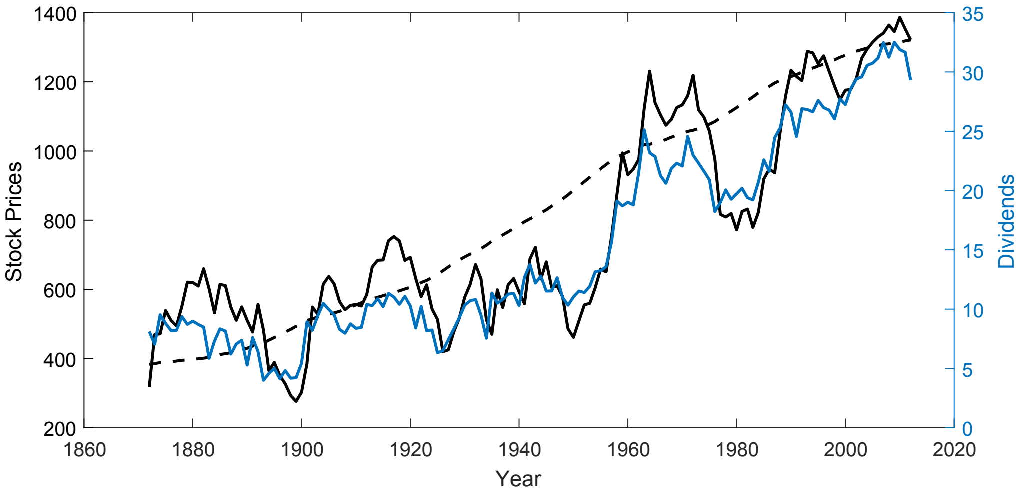
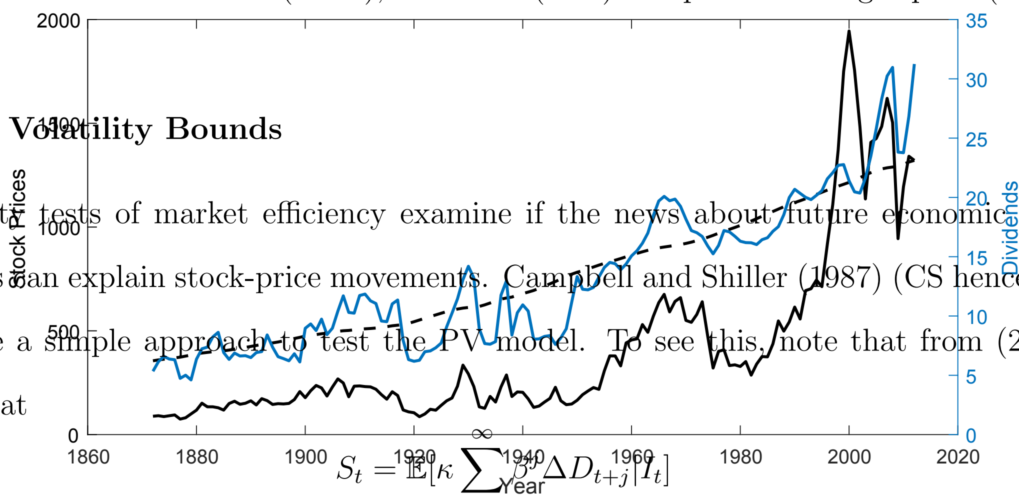


Figure 2: Simulation: Price (black), dividends (blue) and perfect foresight price (dotted).



3.1.2 Volatility Bounds

Volatility tests of market efficiency examine if the news about future economic fundamentals can explain stock-price movements. Campbell and Shiller (1987) (CS henceforth) propose a simple approach to test the PV model. To see this, note that from (2.6) we have that

$$S_t = \mathbb{E}[\kappa \underbrace{\sum_{j=1}^{\infty} \beta^j \Delta D_{t+j}|I_t]}_{S_t^*} \tag{3.2}$$

where $S_t \equiv P_t - \kappa D_{t-1}$ is stationary, since prices and dividends are cointegrated. CS propose to check if

$$\frac{\text{var}(S_t)}{\text{var}(S_t')} = 1,$$

where $S_t' \equiv \mathbb{E}[S_t^*|H_t]$, is the optimal forecast of S_t^* based on the econometrician's information set. In particular, CS obtain S_t' from an unrestricted VAR model. CS find that $\frac{\text{var}(S_t)}{\text{var}(S_t')}$ is considerably larger than one[6] and conclude that: "our evaluation of the present value

[6]Specifically, with 8.2% discount rate $\frac{\text{var}(S_t)}{\text{var}(S_t')} = 67.22$, and with 3.2% discount rate $\frac{\text{var}(S_t)}{\text{var}(S_t')} = 4.786$.

model for stocks indicates that the spread between stock prices and dividends moves too much" (p. 1086). Using a similar approach, Campbell and Shiller (1988a,b) arrive at the same conclusions.

This conclusion is based on the assumption that the econometrician's information set is equivalent to the agent's information set. The assumption that H_t is equivalent to I_t is obviously an strong assumption, but CS argue that under the null hypothesis that (3.2) is true, stock prices should reflect investor's information about discounted value of future dividends. Interestingly, that is not the case here. As we have seen in Section 2.1, an econometrician who uses standard econometrics techniques such as VAR, *unknowingly* conditions on a smaller information set than agents. As a result, the PV model implies that $\frac{\text{var}(S_t)}{\text{var}(S'_t)} \geq 1$.

Proposition 1: Let $S'_t \equiv \mathbb{E}[S^*_t|H_t]$, $S_t \equiv \mathbb{E}[S^*_t|I_t]$ and $H_t \subseteq I_t$. Then the present value relation implies that

$$\text{var}(S'_t) \leq \text{var}(S_t) \tag{3.3}$$

Proof: From the law of iterated projections we have

$$S'_t \equiv \mathbb{E}[S^*_t|H_t] = \mathbb{E}\big[\mathbb{E}[S^*_t|I_t]|H_t\big] = \mathbb{E}[S_t|H_t].$$

The proof is complete upon noticing that $\text{var}\big(\mathbb{E}(x|I_t)\big) \leq \text{var}(x)$ for any random variable x.$\square$

The intuition behind Proposition 1 is simple. S'_t and S_t are two different optimal forecasts of S^*_t, based on two different information sets. Different forecasts decompose the variance of S^*_t into two components, the forecast component and the forecast error component. Specifically

$$S_t = \mathbb{E}[S_t^*|I_t] \qquad \Rightarrow \qquad \mathrm{var}(S_t^*) = \mathrm{var}(S_t) + \mathrm{var}(\zeta_t)$$
$$S_t' = \mathbb{E}[S_t^*|H_t] \qquad \Rightarrow \qquad \mathrm{var}(S_t^*) = \mathrm{var}(S_t') + \mathrm{var}(\zeta_t')$$

Since the variance of the forecast error with less information is at least as large as the variance of the forecast error with more information (i.e., $\mathrm{var}(\zeta_t) \leq \mathrm{var}(\zeta_t')$), the inequality (3.3) must hold.

To summarize, if stock prices are forward-looking, Campbell and Shiller (1987, 1988a,b) bounds are lower bounds on stock prices, not upper bounds. Other volatility tests that do not use VAR techniques, such as Cochrane (1992), do not find excess volatility.

3.2 Cross-Equation Restrictions

A distinguishing characteristic of rational expectations hypothesis is that the parameters describing the stochastic environment that the agents confront appears in the equilibrium solution. Campbell and Shiller (1987) propose a convenient method for characterizing the cross-equation restrictions that the PV relation imposes on the data. These restrictions are frequently rejected by the data, which has been interpreted as evidence against constant discount factors or sometimes against rational expectation hypothesis. In the following, I argue that these rejections could reflect the violation of the underlying assumptions used to drive these cross-equation restrictions.

Present value model implies that

$$S_t = \mathbb{E}[\kappa \sum_{j=1}^{\infty} \beta^j \Delta D_{t+j} | I_t]. \tag{3.4}$$

Assuming that the model has a backward-looking VAR representation

$$\underbrace{\begin{bmatrix} \Delta D_t \\ S_t \end{bmatrix}}_{Z_t} = \underbrace{\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} \Delta D_{t-1} \\ S_{t-1} \end{bmatrix}}_{Z_{t-1}} + \underbrace{\begin{bmatrix} \eta_{1,t} \\ \eta_{2,t} \end{bmatrix}}_{\eta_t},$$

Campbell and Shiller (1987) derive these restrictions by projecting (3.4) onto H_t, and "assuming" that $\mathbb{E}[Z_{t+j}|H_t] = A^j Z_t$

$$\mathbf{e1}' Z_t = \kappa \sum_{j=1}^{\infty} \beta^j \mathbf{e2}' A^j Z_t \tag{3.5}$$

where $\mathbf{e1}' = [0 \;\; 1]$ and $\mathbf{e2}' = [1 \;\; 0]$. Since equation (3.5) holds for all realizations of Z_t, we have

$$\mathbf{e1}' = \kappa \sum_{j=1}^{\infty} \beta^j \mathbf{e2}' A^j = \kappa \mathbf{e2}' \beta A (\mathbf{I} - \beta A)^{-1} \tag{3.6}$$

where the second equality follows by evaluating the infinite sum, which must converge because the elements of Z_t are stationary. Postmultiplying both sides of (3.6) by $(\mathbf{I} - \beta A)$, we have

$$\mathbf{e1}'(\mathbf{I} - \beta A) = \kappa \mathbf{e2}' \beta A. \tag{3.7}$$

When these restrictions are applied to PV asset pricing models, the finding is almost always a resounding rejection. Note that these restrictions are derived by imposing the arbitrary assumption that $\mathbb{E}[Z_{t+j}|H_t] = A^j Z_t$. However, in our framework $\mathbb{E}[Z_{t+j}|H_t] \neq A^j Z_t$, since the matrix A has at least one root inside the unit circle. Therefore, the rejections of cross-equation restrictions could reflect the rejections of the underlying hypothesis used to derive (3.7). This proves the following proposition.

Assumption 1. $\{\eta_t\}$ are *iid* and mutually independent, strictly stationery process with a non-Gaussian distribution such that $(a + 1)$st moment finite with $(a + 1)$st cumulant nonzero for some $a \geq 2$.

Proposition 2: Under assumption 1, the best predictor of a noncausal model is nonlinear. Therefore, standard cross-equation restriction tests that exclude forward-looking representations, can produce spurious rejections.

4 Is the US Stock Market Efficiently React to News?

From the Analytical example in Section 2 we saw that market efficacy imposes an upper bound on the contribution of noise to price volatility; that is, noise explains less than half of price volatility. In this section, I propose a simple procedure to test this implication. The first step is to define noise. The basic idea of the decomposition is most easily explained in the context of a general Euler equation:

$$P_t = \mathbb{E}\big[M_{t+1}(P_{t+1} + D_t)|I_t\big].$$

Iterating forward and imposing the transversality condition, we obtain the PV model with time-varying discount rates

$$P_t = \mathbb{E}\left[\sum_{j=1}^{\infty}\Big(\prod_{i=1}^{j} M_{t+1}\Big) D_{t+j}|I_t\right]. \tag{4.1}$$

Since prices and dividends are not stationary, population moments can not be estimated from sample counterparts. However, if dividend growth, $\frac{D_t}{D_{t-1}} = G_t$, and discount rates are stationary, the PV model implies that price-dividend ratio is stationary.[7] Thus, we can write (4.1) as

$$\frac{P_t}{D_t} = \mathbb{E}\left[\sum_{j=1}^{\infty}\prod_{i=1}^{j} X_{t+i}|I_t\right], \tag{4.2}$$

[7] For the formal proof see (Cochrane, 1994).

where $X_t = M_t G_t$ is the discounted dividend growth rate, which can be obtained directly from the data. Notice that X_t captures all the economic fundamentals that is important for price-dividend ratio. According to the PV model, a high price-dividend ratio implies that either future dividend growth is high, or future returns is low, or some combination of the two. For the discount rates, I use the realized returns, $R_t = \frac{P_t + D_t}{P_{t-1}}$. Notice that we do not require any specific assumption on the time series of dividends or returns, and therefore is not subject to the joint hypothesis problem. Moreover, since we include all the relevant information -according to the present value model- for the price fluctuation, the test is robust to the triple hypothesis problem.

Following the discussion in Section 2, I define noise as the component of price that is orthogonal to the future economic fundamentals, X_t. Therefore, I approximate the noise component by the residuals of the following linear projection,

$$\frac{P_t}{D_t} = \gamma_0 + \gamma_1 \sum_{j=1}^{K} \prod_{i=1}^{j} X_{t+i} + e_t \tag{4.3}$$

where e_t is orthogonal to the right hand side by construction, and the fitted value of the regression approximates the fundamental component.[8]

What should we expect? If investors have no information about the future economic fundamentals, then PV model implies that the price-dividend ratio is constant, which it is not. On the other hand, if investors have perfect knowledge of future economic fundamentals, the price-dividend ratio lies in the space spanned by the future X_t's, which implies that $R^2 = 1$. However, a more realistic assumption is that investors have some

[8]Note that regression (4.3) is different from the long-run predictability regressions

$$\sum_{j=1}^{K} Y_{t+j} = \gamma_0 + \gamma_1 \frac{P_t}{D_t} + e_t,$$

where $Y_t = R_t$ or ΔD_t. The literature almost exclusively focused on the relative importance of dividend growth and discount factors. In contrast, I am only interested in the comovement of the price-dividend ratio with future economic fundamentals. However, if we define $Y_t = M_t G_t$, the R^2 would be the same as the R^2 from the regression (4.3).

noisy information about the future economic fundamentals, the $R^2 \leq 1$. Then market efficiency then implies that noise can not explain more than 50% of price-dividend fluctuations. Therefore, if we find that $R^2 \leq 0.5$ we can reject the hypothesis that markets were reacting efficiently to the news.

Table (1) presents summary statistics and the results, where I set $K = 10$. The results are not particularly sensitive to the choice of K. As predicted by the PV model (4.2), price-dividend ratio is positively correlated with the fitted value of regression (4.3), shown in the first row. The $R^2 = 0.44$ is not a good news for market efficiency supporters. However, one must be careful concluding that markets are inefficient. *First*, we should not expect that the parameters did not change through more than 140 years. For instance, we have extensive empirical evidence of dividends smoothing in the postwar period than before.[9] Table (1) also presents empirical results both for two subsamples 1871-1945 and 1946-2012. Consistent with market efficiency and rationality, I find that $R^2 = 0.67$ and $R^2 = 0.51$, respectively. *Second*, these results are based on the assumption that investors' optimal forecast is linear. While the linear specification provides important insights, there is no theoretical reason why the optimal forecast must be linear. A tractable alternative is to formulate the decomposition problem as a linear-quadratic regression,

$$\frac{P_t}{D_t} = \gamma_0 + \gamma_1 \sum_{j=1}^{10} \prod_{i=1}^{j} X_{t+j} + \gamma_2 \sum_{j=1}^{10} \prod_{i=1}^{j} X_{t+j}^2 + e_t \tag{4.4}$$

The results are reported in Panel B, Table (1). Again, the R^2 is grater than 0.5 in all periods, which is consistent with market efficiency. Including cubic terms slightly improves the R^2.

Thus, a large component of price-dividend ratio fluctuations is due to noise. The interesting question naturally arises is about other source of variation. According to the PV model, high price-dividend ratio implies that either future dividend growth must

[9]See, for instance, Chen et al. (2012).

Table 1: Testing For Market Efficiency

Sample Period	γ_0	γ_1	γ_2	R^2
Panel A: linear specifications				
1871 – 2014	-10.4 (3.5)	4.4 (0.4)	-	**0.44**
1871 – 1945	0.01 (2.5)	1.6 (0.2)	-	**0.69**
1946 – 2014	-6.5 (5.1)	4.7 (0.6)	-	**0.51**
Panel B: linear-quadratic specifications				
1871 – 2014	40.8 (9.0)	-14.9 (3.2)	6.9 (1.14)	**0.56**
1871 – 1945	-22.7 (7.2)	12.4 (3.1)	-3.8 (1.2)	**0.72**
1946 – 2014	27.3 (14.4)	-8.18 (5.2)	4.49 (1.8)	**0.54**

Notes: The regression equations considered are (4.3) and (4.4), respectively. The numbers in the parenthesis are the standard errors.

be high, or future discount rates must be low, or both. To evaluate the contribution of each component, I repeat the exercise but fix the returns or the dividends growth to their respective mean when defining Y_t. First, I examine dividend growth. In the regressions of Table (2), shown in the second row, price-dividend ratio does not comove with future dividend growth. The coefficient of the dividend growth is not significant and the correlation is almost *zero*. Next, I examine the long-horizon returns. It turns out that the price-dividend ratio comoves more strongly with the fitted part $\mathrm{corr}(\frac{P_t}{D_t}, \frac{\hat{P}_t}{D_t}) = 0.39$, and the comovement increases in the post WWII period, $R^2 = 0.58$. Therefore, consistent with previous empirical findings,[10] it appears that variation in price-dividend is mostly driven by movements in discount rates, instead of movements in expected dividend growth.[11]

Overall, this procedure does a fairly good job explaining the behavior of stock prices. The historical decompositions of price-dividend ratios are presented in Figure (3). From these figures we see that the late 1990s and early 2000s market was overvalued. Shiller refers to this period as *millennium bubble* or the *dot-com bubble*. He uses Cyclically Adjusted Price-Earnings (CAPE) ratio, originally defined by Campbell and Shiller (1998), as an informal measure to characterize bubbles. CAPE ratio equals the S&P Composite Index, divided by the ten-year moving average of real earnings on the index to smooth out the effects of economic cycles:

$$\mathrm{CAPE}_t = \frac{P_t}{\left[(\mathrm{EARN}_t + \mathrm{EARN}_{t-1} + \cdots + \mathrm{EARN}_{t-10})/10\right]}.$$

Shiller date stamps the dot-com bubble in the 1982-2000 period. My proposed method dates the beginning at 1987. The difference could be due to two reasons. First, CAPE is a backward measure which does not include future information. Second, date stamping

[10]See for example Cochrane (2011) and the references therein.

[11]Dividend smoothing might have an effect in obtaining such a low correlation. See for example Chen et al. (2012).

Table 2: Historical Decomposition of Price-Dividend Ratio

X_t	γ_0	γ_1	corr$(\frac{P_t}{D_t}, \frac{\hat{P}_t}{D_t})$	R^2
		Sample 1871-2014		
$\bar{M}G_t$	19.8* (6.5)	0.7 (0.9)	**0.06**	**0.01**
$M_t\bar{G}$	8.9* (3.4)	1.8* (0.4)	**0.39**	**0.15**
		Sample 1871-1945		
$\bar{M}G_t$	18.4* (2.0)	0.16 (0.3)	**0.06**	**0.01**
$M_t\bar{G}$	15.8* (1.4)	0.43* (0.2)	**0.31**	**0.10**
		Sample 1946-2014		
$\bar{M}G_t$	21.9 (18.5)	1.3 (2.5)	**0.06**	**0.01**
$M_t\bar{G}$	3.3* (0.61)	1.6* (0.2)	**0.58**	**0.34**

Notes: The regression equation is $\frac{P_t}{D_t} = \gamma_0 + \gamma_1 \sum_{j=1}^{10} \prod_{i=1}^{j} X_{t+j} + e_t$. $\frac{\hat{P}_t}{D_t}$ is the fitted value of the regression. $\bar{G}$ denotes the mean of dividends growth, $G_t = \frac{D_t}{D_{t-1}}$, and $\bar{M}$ denotes the mean of realized returns. The fitted value, $\frac{\hat{P}_t}{D_t} = \hat{\gamma}_0 + \hat{\gamma}_1 \sum_{j=1}^{10} \prod_{i=1}^{j} X_{t+j}$, approximates the fundamental component. The numbers in the parenthesis are the standard errors. Data are annual, 1871-2012.

according to CAPE is arbitrary, since the fundamental value is not defined explicitly.

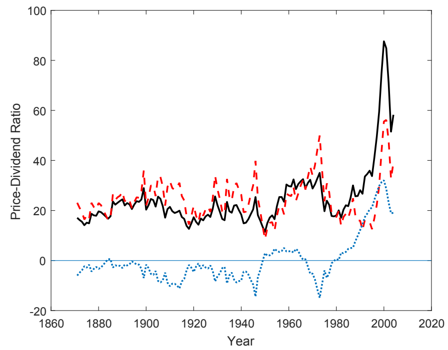
Consistent with Irving Fischer's observation, I find no evidence that the market was overvalued before the Wall Street Crash of 1929. Using data on productive capital on stocks and tax rates to estimate the fundamental value, McGrattan and Prescott (2001) also find that stock prices were undervalued, even at their peak.[12]

Figure (3) shows that stock market can remain undervalued for extended period of time. Negative bubbles were present in at least two periods: 1938-1949, and 1965-1980. Modigliani and Cohn (1979) hypothesize that the stock market suffers from *money illusion*, discounting real dividends using nominal discount rates. An implication of such irrational behavior is that when inflation is high (as it was the case during 1970s), the

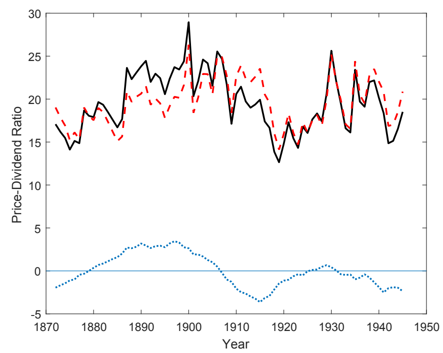
[12] See also Donaldson and Kamstra (1996).

Figure 3: Historical Decomposition of Price-Dividend Ratio

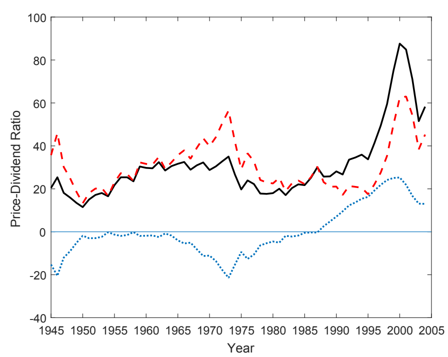
(a) Sample 1871-2012



(b) Sample 1871-1945

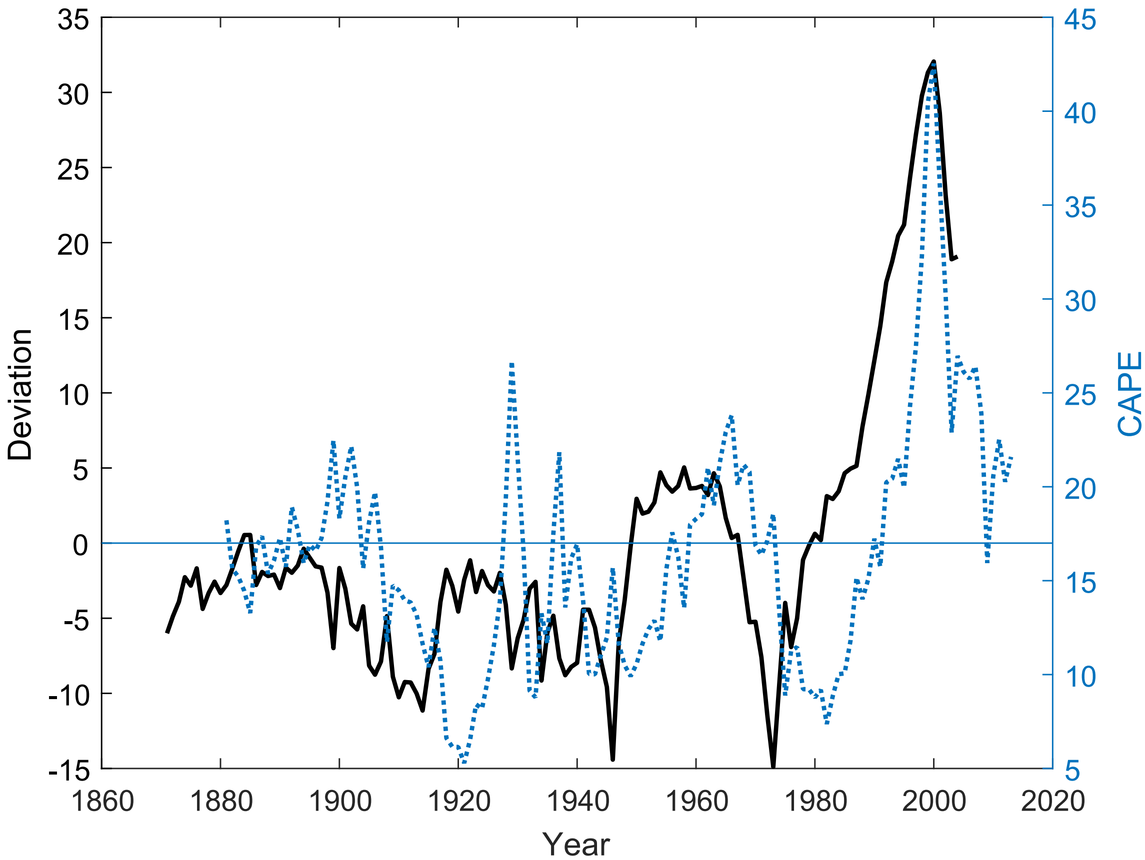


(c) Sample 1946-2014



Notes: Price-dividend ratio (solid, black), estimated fundamental component (dashed, red) given by the fitted value of (4.3), and estimated noise component (dotted, blue) which is given by the residuals of the (4.3). Annual data, 1871-2012.

Figure 4: Comparison of CAPE and Fundamental Component



Notes: Percentage deviation of fundamental component from price-dividend ratio (solid, black), Cyclical adjusted price-earnings ratio (dotted, blue). Annual data, 1871-2012.

stock market is undervalued. Fama (1981) gives a rational interpretation of this phenomenon: high inflation signals a decline in future economic activities, and the stock market rationally reflects the decline into prices.

5 Conclusions

I have shown that news about future economic fundamentals can create rational expectations equilibria with non-causal representations that pose substantial challenges to the estimation and inference of conventional econometric analyses. Many of the estimation procedures in econometrics packages can not adequately estimate the non-causal processes. An outside econometrician who uses standard econometrics techniques, such as VARs, will find patterns that are different from those perceived by rational agents. This includes excess volatility, violations of cross-equations restrictions, volatility clustering and rejections of forecast rationality tests. In Section 2, I present a modified present value model to describe the nature of the problem. The key insight is the role of noisy information about the future economic fundamentals. Section 3 demonstrates that failing to model non-causality can produce important inferential errors.

In the empirical section I propose a new method for testing market efficiency, which is robust to the joint hypothesis problem. I show that market efficiency imposes restrictions on the contribution of noise to the price volatility. To perform the test, we need a measure of the noise in the market. Therefore, I propose a new procedure to decompose stock prices into a fundamental component and a noise component. Applying my procedure to the US data, I find (1) US stock market has been efficiently reacting to news; (2) variation in price-dividend is mostly driven by movements in discount rates, instead of movements in expected dividend growth. Moreover, my estimates show that extensive evidence of stock market overvaluation and undervaluation. It turns out that the observed variation in price-dividend ratio is mainly due to news about long-horizon returns, and none is due to news about dividend-growth. I find that the US stock market was undervalued during the 1970s and overvalued during the 1990s. Interestingly, I find no evidence the market was overvalued before the Wall Street Crash of 1929.

Appendix A

In this section we generalize the simple model of section 2. The assumptions on the model specification and expectations is as before, but I extend the information set of the agents. As before, I assume that agents observe the complete history of prices and dividends. However, at the beginning of each period and right before prices are determined, agents observe some noisy signals about the entire future dividend changes. Following the notation that $s_{t|t+j}$ denotes the news at time t about time $t+j$ dividends, we have

$$s_{t|t+k} = \epsilon_{t+k} + \sum_{i=0}^{k} \nu_{t+i} \quad \text{for all } k \geq 0, \tag{A.1}$$

where $\nu_t \sim iid(0, \sigma_\nu^2)$ denotes the noise which is orthogonal to ϵ_t at all leads and lags. Thus, the agents' information set at the begging of period t (say I_t), encompass the current and past signals, as well as the history of observed prices and dividends.

Therefore, $\mathbb{E}[D_{t+j}|I_t]$, for $j = 0, 1, 2, \cdots$, is simply the linear projection of D_{t+j} on I_t, which implies that

$$\begin{aligned} \mathbb{E}[D_{t+j}|I_t] &= D_{t-1} + \sum_{k=0}^{j} \frac{1}{1+(k+1)\gamma} s_{t|t+k}, \\ &= D_{t-1} + \sum_{k=0}^{j} \frac{1}{1+(k+1)\gamma} (\epsilon_{t+k} + \vartheta_{t+k}),\ j = 0, 1, 2, \cdots \end{aligned} \tag{A.2}$$

where $\gamma = \frac{\sigma_\nu^2}{\sigma_\epsilon^2}$ and $\vartheta_{t+k} = \sum_{i=0}^{k} \nu_{t+i}$ denotes the aggregate noise. From (A.2) we see that agents increasingly discount the signals further into the future. For instance, a signal about D_t is discounted by $\frac{1}{1+\gamma}$, but the signal about D_{t+1} is discounted by $\frac{1}{1+2\gamma}$, and so forth. Intuitively, the signals about the distant future are less accurate and investors do not react to them the same way they react to accurate signals.

We can now find equilibrium stock prices. Solving (2.1) forward, together with the transversality condition we obtain a simple Present Value (PV) model with constant discount factor

$$P_t = \underbrace{\sum_{j=0}^{\infty} \beta^{j+1} \mathbb{E}(D_{t+j}|I_t)}_{fundamental\ value}. \tag{A.3}$$

Substituting (A.2) into (A.3), the generalized form of equilibrium price is given by

$$\begin{aligned} P_t &= \kappa D_{t-1} + \kappa \sum_{j=0}^{\infty} \frac{\beta^j}{1+(j+1)\gamma} s_{t|t+j} \\ &= \underbrace{\kappa D_{t-1} + \kappa \sum_{j=0}^{\infty} \frac{\beta^j}{1+(j+1)\gamma} \epsilon_{t+j}}_{fundamental\ component} + \underbrace{\kappa \sum_{j=0}^{\infty} \frac{\beta^j}{1+(j+1)\gamma} \vartheta_{t+j}}_{noise\ component}, \end{aligned} \tag{A.4}$$

where $\kappa = \frac{\beta}{1-\beta}$.

Appendix B

I first prove Lemma 1, which is based on Rosenblatt (2000). I use Lemma 1 to prove the univariate case of Proposition 2.1, and then show that under Assumption 1 the multivariate case can be reduced to the univariate case.

Lemma 1: Consider a univariate noncausal AR(p), that is

$$\phi(z)x_t = \eta_t,$$

where $\phi(z)$ has at least one root inside the unite circle. Under Assumption 1, Theorem 5.4.1 and Corollary 5.4.2 of Rosenblatt (2000) implies that the best predictor of x_t condition on its past is nonlinear, which then implies that the Wold innovations, $\tilde{\eta}_t$, are non-mds.

Proof of Proposition 2.1: The proof is similar to the Corollary 2.1 in Hamidi Sahneh (2015).□

References

Beaudry, P. and Portier, F., 2006. Stock prices, news, and economic fluctuations. *The American economic review*, 96(4):1293–1307.

Black, F., 1986. Noise. *The journal of finance*, 41(3):528–543.

Breidt, F. J., Davis, R. A., and Trindade, A. A., 2001. Least absolute deviation estimation for all-pass time series models. *Annals of statistics*, pages 919–946.

Brockwell, P. J. and Davis, R. A., 2013. *Time series: theory and methods*. Springer Science & Business Media.

Brunnermeier, M. K., 2009. Bubbles: Entry in new palgrave dictionary of economics. *ed. Steven Durlauf and Lawrence Blume*.

Campbell, J. Y. and Shiller, R. J., 1987. Cointegration and tests of present value models. *The Journal of Political Economy*, pages 1062–1088.

Campbell, J. Y. and Shiller, R. J., 1988a. The dividend-price ratio and expectations of future dividends and discount factors. *Review of financial studies*, 1(3):195–228.

Campbell, J. Y. and Shiller, R. J., 1988b. Stock prices, earnings, and expected dividends. *The Journal of Finance*, 43(3):661–676.

Campbell, J. Y. and Shiller, R. J., 1998. Valuation ratios and the long-run stock market outlook. *The Journal of Portfolio Management*, 24(2):11–26.

Chen, L., Da, Z., and Priestley, R., 2012. Dividend smoothing and predictability. *Management science*, 58(10):1834–1853.

Cochrane, J. H., 1992. Explaining the variance of price–dividend ratios. *Review of Financial Studies*, 5(2):243–280.

Cochrane, J. H., 1994. Permanent and transitory components of gnp and stock prices. *The Quarterly Journal of Economics*, pages 241–265.

Cochrane, J. H., 2011. Presidential address: Discount rates. *The Journal of Finance*, 66 (4):1047–1108.

Davis, R. and Song, L., 2012. Noncausal vector ar processes with application to economic time series. *DP Columbia University*.

Diba, B. T. and Grossman, H. I., 1988. Explosive rational bubbles in stock prices? *The American Economic Review*, pages 520–530.

Donaldson, R. G. and Kamstra, M., 1996. A new dividend forecasting procedure that rejects bubbles in asset prices: The case of 1929s stock crash. *Review of Financial Studies*, 9(2):333–383.

Fama, E. F., 1970. Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, 25(2):383–417.

Fama, E. F., 1981. Stock returns, real activity, inflation, and money. *The American Economic Review*, 71(4):545–565.

Farmer, J. D. and Geanakoplos, J., 2009. The virtues and vices of equilibrium and the future of financial economics. *Complexity*, 14(3):11–38.

Forni, M., Gambetti, L., Lippi, M., and Sala, L., 2016. Noise bubbles. *The Economic Journal*.

Gouriéroux, C. and Zakoïan, J.-M., 2016. Local explosion modelling by noncausal process. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. ISSN 1467-9868.

Grossman, S. J. and Shiller, R. J., 1981. The determinants of the variability of stock market prices. *The American Economic Review*, 71(2):222–227.

Hamidi Sahneh, M., 2015. Are the shocks obtained from svar fundamental?

Hamilton, J. D., 1994. *Time series analysis*. Princeton university press Princeton.

Hansen, L. P., Sargent, T. J., et al. Exact linear rational expectations models: Specification and estimation. Technical report, 1981.

Hong, Y., 1999. Hypothesis testing in time series via the empirical characteristic function: A generalized spectral density approach. *Journal of the American Statistical Association*, 94(448):1201–1220.

Hong, Y. and Lee, Y.-J., 2005. Generalized spectral tests for conditional mean models in time series with conditional heteroscedasticity of unknown form. *The Review of Economic Studies*, 72(2):499–541.

Jondeau, E. and Rockinger, M., 2003. Conditional volatility, skewness, and kurtosis: existence, persistence, and comovements. *Journal of Economic Dynamics and Control*, 27(10):1699–1737.

Kasa, K., Walker, T. B., and Whiteman, C. H., 2014. Heterogeneous beliefs and tests of present value models. *The Review of Economic Studies*, 81(3):1137–1163.

Kleidon, A. W., 1986. Variance bounds tests and stock price valuation models. *The Journal of Political Economy*, pages 953–1001.

Lanne, M. and Saikkonen, P., 2013. Noncausal vector autoregression. *Econometric Theory*, 29(03):447–481.

Leeper, E. M., Walker, T. B., and Yang, S.-C. S., 2013. Fiscal foresight and information flows. *Econometrica*, 81(3):1115–1145.

Lii, K.-S. and Rosenblatt, M., 1982. Deconvolution and estimation of transfer function phase and coefficients for nongaussian linear processes. *The Annals of Statistics*, pages 1195–1208.

Lippi, M. and Reichlin, L., 1994. Var analysis, nonfundamental representations, blaschke matrices. *Journal of Econometrics*, 63(1):307–325.

Malkiel, B. G. The efficient-market hypothesis and the financial crisis. In *Rethinking Finance: Perspectives on the Crisis (Proceedings of a Conference). Russel Sage Foundation*, 2011.

McGrattan, E. R. and Prescott, E. C. The stock market crash of 1929: Irving fisher was right! Technical report, National Bureau of Economic Research, 2001.

Modigliani, F. and Cohn, R. A., 1979. Inflation, rational valuation and the market. *Financial Analysts Journal*, 35(2):24–44.

Phillips, P. C., Wu, Y., and Yu, J., 2011. Explosive behavior in the 1990s nasdaq: When did exuberance escalate asset values? *International economic review*, 52(1):201–226.

Rosenblatt, M., 2000. *Gaussian and non-Gaussian linear time series and random fields.* Springer.

Shiller, R. J., 1981. Do stock prices move too much to be justified by subsequent changes in dividends ? *The American Economic Review*, 71(3):421–436.

Shiller, R. J., 2014. Speculative asset prices. *The American Economic Review*, 104(6): 1486.